

NBER WORKING PAPER SERIES

OPTIMAL DEFAULTS AND ACTIVE DECISIONS

Gabriel D. Carroll
James J. Choi
David Laibson
Brigitte Madrian
Andrew Metrick

Working Paper 11074
<http://www.nber.org/papers/w11074>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2005

We thank Hewitt Associates for their help in providing the data. We are particularly grateful to Lori Lucas, Jim McGhee, Scott Peterson, Deborah Bonus, and Yan Xu, some of our many contacts at Hewitt. We also appreciate many helpful conversations with Harlan Fleece and the exceptional research assistance of Nelson Uhan, Jared Gross, John Friedman, Hongyi Li, Laura Serban, Keith Ericson, Ben Grossman, and Fuad Faridi. Richard Thaler, Mark Iwry, Shlomo Benartzi, Annamaria Lusardi, and seminar participants at the University of Chicago, Harvard, MIT, Berkeley, Cornell, USC, LSE, Wharton, Dartmouth, Stanford, the New York Federal Reserve, the Russell Sage Foundation, and the NBER have provided much useful feedback. Choi acknowledges financial support from a National Science Foundation Graduate Research Fellowship and the Mustard Seed Foundation. Choi, Laibson, and Madrian acknowledge individual and collective financial support from the National Institute on Aging (grants R01-AG-16605, R29-AG-013020, and T32-AG00186). Laibson also acknowledges financial support from the Sloan Foundation.

© 2005 by Gabriel D. Carroll, James J. Choi, David Laibson, Brigitte Madrian, and Andrew Metrick. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Optimal Defaults and Active Decisions

Gabriel D. Carroll, James J. Choi, David Laibson, Brigitte Madrian, and Andrew Metrick

NBER Working Paper No. 11074

January 2005, Revised October 2007

JEL No. D0,E21,G23

ABSTRACT

Defaults can have a dramatic influence on consumer decisions. We identify an overlooked but practical alternative to defaults: requiring individuals to make an explicit choice for themselves. We study such "active decisions" in the context of 401(k) saving. We find that compelling new hires to make active decisions about 401(k) enrollment raises the initial fraction that enroll by 28 percentage points relative to a standard opt-in enrollment procedure, producing a savings distribution three months after hire that would take three years to achieve under standard enrollment. We also present a model of 401(k) enrollment and characterize the optimal enrollment regime. Active decisions are optimal when consumers have a strong propensity to procrastinate and savings preferences that are highly heterogeneous. Naive beliefs about future time-inconsistency strengthen the normative appeal of the active-decision enrollment regime. However, financial illiteracy favors default enrollment over active decision enrollment.

Gabriel D. Carroll
MIT Department of Economics
50 Memorial Drive
Cambridge, MA 02142
gdc@mit.edu

James J. Choi
Yale School of Management
135 Prospect Street
P.O. Box 208200
New Haven, CT 06520-8200
and NBER
james.choi@yale.edu

David Laibson
Department of Economics
Littauer M-14
Harvard University
Cambridge, MA 02138
and NBER
dlaibson@harvard.edu

Brigitte Madrian
John F. Kennedy School of Government
Harvard University
79 JFK Street
Cambridge, MA 02138
and NBER
Brigitte_Madrian@Harvard.edu

Andrew Metrick
The Wharton School
University of Pennsylvania
3620 Locust Walk
Philadelphia, PA 19104
and NBER
metrick@wharton.upenn.edu

Ex ante, it might seem that a default should not matter if agents believe it is arbitrarily chosen and if opting out of the default is easy. In practice, defaults—even bad defaults—tend to be sticky; individuals often fail to opt out.¹ This perverse property of defaults has been documented in a wide range of settings: participation in employer-sponsored savings plans (Madrian and Shea, 2001; Choi, Laibson, Madrian and Metrick, 2004), organ donation rates (Johnson and Goldstein, 2003; Abadie and Gay, 2004), car insurance plan choices (Johnson et al., 1993), car option purchases (Park, Yun and MacInnis, 2000), and consent to receive e-mail marketing (Johnson, Bellman, and Lohse, 2002).

In light of this inertia, defaults may be socially optimal when agents have a shared optimum and the default leads them to it. But even a well-chosen default may be undesirable if agents have heterogeneous needs. For example, in a firm whose workforce includes young, cash-strapped single parents and older employees who need to quickly build a retirement nest egg, one 401(k) savings rate isn't right for everyone.

Given that defaults can exert a tremendous impact on observed outcomes, an important question is, “which default is optimal?” We consider the answer to this question in a specific context: participation in employer-sponsored 401(k) savings plans. Economists have studied two different savings plan enrollment regimes. Under “standard enrollment,” employees are by default not enrolled and can choose to opt into the plan. Under “automatic enrollment,” employees are by default enrolled at a pre-determined contribution rate and can choose to opt out of the plan or out of the default contribution rate. There is also a third, overlooked alternative: require individuals to make an explicit choice for themselves. In this “active-decision” regime, there is *no* default to fall back on.

Active decisions are an intriguing, though imperfect, alternative to defaults. On the positive side, an active-decision mechanism avoids the biases introduced by defaults because it does not corral agents into a single default choice. The active-decision mechanism encourages agents to think about an important decision and thereby avoid procrastinating. On the negative side, an active decision compels agents to struggle with a potentially time-consuming

¹For example, about three-quarters of 401(k) participants in firms with automatic enrollment retain both the default contribution rate and the default asset allocation (Madrian and Shea, 2001; Choi et al. 2002, 2004). These “choices” are puzzling because most companies with automatic enrollment have had very conservative defaults; a typical firm might have a default contribution rate of 3% of income, even though contributions up to 6% of income garner matching contributions from the employer.

decision—which they may not feel qualified to make. Some individuals would welcome a benign third party who is willing to make that decision for them. And some social engineers might prefer a default that aggressively encourages some social goal, like organ donation or retirement saving.²

The current paper lays the groundwork for a debate about active decisions by describing how an active-decision 401(k) enrollment regime worked at one large firm. In addition, we present a model that provides a formal framework for evaluating the relative efficacy of different enrollment mechanisms, including defaults and active decisions.

Our empirical analysis exploits a natural experiment in the 401(k) enrollment regime of a large firm that switched from using active decisions to a standard-enrollment regime. This change in 401(k) enrollment procedures occurred unintentionally as a by-product of the transition from a paper-and-pencil administrative system to a phone-based administrative system. The firm did not anticipate that the transition to a phone-based system with a default of non-enrollment would transform the psychology of 401(k) participation. Rather, the change in administrative systems was motivated solely by the convenience and efficiency of phone-based enrollment. The loss of active-decision effects was a collateral consequence of that transition.

We find that active decisions raise the initial fraction of employees enrolled by 28 percentage points relative to what is obtained with a standard default of non-enrollment. Active decisions raise average savings rates and accumulated balances by accelerating decision-making. We show that conditional on demographics, employees under an active-decision regime will on average *immediately* choose a savings rate similar to what would otherwise take up to 30 months to attain under standard enrollment. Because the typical worker will change jobs several times before retirement, accelerating the 401(k) savings decision by more than two years at the beginning of each job transition can have a large impact on accumulated wealth at retirement.

We also present a model of the enrollment process. In this model, defaults matter for two key reasons. First, the opportunity cost of time is stochastic, creating an option value to waiting for a low-cost time to take action. Second, employees with present-biased preferences

²See Hurst (2006) and Warshawsky and Ameriks (2000) for evidence that many U.S. households are undersaving. However, this is an open question with research on both sides of the debate. See Aguiar and Hurst (2005) and Engen, Gale, and Uccello (1999) for evidence that the lifecycle model with liquidity constraints matches U.S. data.

(Phelps and Pollak, 1968; Laibson, 1997; O’Donoghue and Rabin 1999b) may procrastinate in their decision to opt out of the default. We derive conditions for the optimal enrollment regime. Active decisions are socially optimal when consumers have highly heterogeneous savings preferences and a strong propensity to procrastinate.

The rest of this paper proceeds as follows. Section 1 describes the details of the two 401(k) enrollment regimes at the company we study. Section 2 describes our data. Section 3 compares the 401(k) savings decisions of employees hired under the active-decision regime to those hired under the standard-enrollment regime. Section 4 presents a model of procrastination for time-inconsistent agents with rational expectations, uses this model to derive the socially optimal enrollment mechanism for such agents, and briefly considers the case of agents with naive expectations about their future time-inconsistency. We also discuss the optimal amount of time agents should be given before a decision deadline binds. Section 5 discusses the key implications of the model. Section 6 concludes and briefly discusses the implementation of active-decision mechanisms.

1 The Natural Experiment

We use employee-level data from a publicly traded Fortune 500 company in the financial services industry. In December 1999, this firm had offices in all 50 states, as well as the District of Columbia and Puerto Rico. This paper considers the 401(k) savings decisions of employees at the firm from January 1997 through December 2001.

Until November 1997, all newly hired full-time employees at the firm were *required* to submit a form within 30 days of their hire date stating their 401(k) participation preferences, *regardless* of whether they wished to enroll or not. Although there was no tangible penalty for non-compliance, human-resource officers at this firm report that only 5% of employees did not return the form.³ We believe that this high compliance rate arose because the form was part of a packet that included other forms that are legally required to be completed (e.g., employment-eligibility verification forms, tax-withholding forms). Moreover, employees who did not return the form were reminded to do so by the human-resources department.

Employees who declined to participate in the 401(k) plan during this initial enrollment period could not subsequently enroll in the plan until the beginning (January 1) of succeeding

³A failure to return the form was ultimately treated as a decision not to enroll in the 401(k).

calendar years. Later in the paper, we will show that this delay did not drive the active-decision effects that we document.

At the beginning of November 1997, the company switched from a paper-based 401(k) enrollment system to a telephone-based system. Employees hired after this change no longer received a 401(k) enrollment form when hired. Instead, they were given a toll-free phone number to call if and when they wished to enroll in the 401(k) plan. We call this new system the “standard-enrollment” regime because its default of non-participation is what has been adopted by most companies. The telephone-based system also allowed employees to enroll on a daily basis, rather than only at the beginning of each calendar year as had previously been the case. This change applied not only to employees hired after November 1997, but to all employees working at the company.

A number of other 401(k) plan features also changed at the same time (November, 1997). We believe that these additional changes made 401(k) participation more attractive, so our estimates of the active-decision effect are a lower bound on the true effect. These other changes include a switch from monthly to daily account valuation, the introduction of 401(k) loans, the addition of two new funds as well as employer stock to the 401(k) investment portfolio,⁴ and a switch from annual to quarterly 401(k) statements. Table 1 summarizes the 401(k) plan rules before and after the November 1997 plan changes.

2 The Data

We have two types of employee data. The first dataset is a series of cross-sections at year-ends 1998, 1999, 2000, and 2001. Each cross-section contains demographic information for everybody employed by the company at the time, including birth date, hire date, gender, marital status, state of residence, and salary. For 401(k) plan participants, each cross-section also contains the date of enrollment and year-end information on balances, asset allocation, and the terms of any outstanding 401(k) loans. The second dataset is a longitudinal history of every individual transaction in the plan from September 1997 through April 2002: savings-rate elections, asset-allocation elections for contributions, trades among funds, loan-based withdrawals and repayments, financial-hardship withdrawals, retirement withdrawals, and

⁴Prior to November 1997, employer stock was available as an investment option only for match balances and contributions made with after-tax money.

rollovers.

To analyze the impact of active decisions on savings outcomes, we compare the behavior of two employee groups: employees hired between January 1, 1997 and July 31, 1997 under the active-decision regime,⁵ and employees hired between January 1, 1998 and July 31, 1998 under the standard-enrollment regime. We refer to the first group as the “active-decision cohort” and the second group as the “standard-enrollment cohort.”

The active-decision cohort is first observed in our cross-sectional data in December 1998, 18 to 24 months after hire, and in the longitudinal data starting in September 1997, 3 to 9 months after hire. The longitudinal data only contain 401(k) participants. The standard-enrollment cohort is also observed in our cross-sectional data starting in December 1998, but this is only 6 to 12 months after their hire date. In the longitudinal data, 401(k) participants from this cohort are observed as soon as they enroll.

Since 401(k) participants are less likely to subsequently leave their employer,⁶ our data structure introduces selection effects that are stronger for the active-decision cohort than the standard-enrollment cohort. To equalize the sample selectivity of the active-decision and standard-enrollment cohorts, we restrict both cohorts to employees who were still at the company in December of the year after their hire year. We have no reason to believe that the turnover rates of employees from these two cohorts were different over these time horizons. The economic environment faced by these two groups of employees was similar until the start of the 2001 recession. In addition, company officials reported no material changes in hiring or employment practices during this period.

Table 2 presents demographic statistics on the active-decision and standard-enrollment cohorts at the end of December in the year *after* they were hired. The cohorts are similar in age, gender composition, income, and geographical distribution. The dimension along which they differ most is marital status, and even here the differences are not large: 56.0% of the active-decision cohort is married, while this is true for only 49.7% of the standard-enrollment cohort. The third column of Table 2 shows that the new-hire cohorts are different from

⁵We exclude employees hired prior to January 1, 1997 because the company made two plan changes that took effect on January 1, 1997. First, the company eliminated a one-year service requirement for 401(k) eligibility. Second, the company changed the structure of its 401(k) match. Although active decisions were used until the end of October 1997, we do not include employees hired from August through October to avoid any confounds produced by the transition to standard enrollment. For example, an enrollment blackout was implemented for several weeks during the transition.

⁶See Even and MacPherson (2005).

employees at the company overall. As expected, the new-hire cohorts are younger, less likely to be married, and paid less on average. The last column reports statistics from the Current Population Survey, providing a comparison between the company’s employees and the total U.S. workforce. The company has a relatively high fraction of female employees, probably because it is in the service sector. Employees at the company also have relatively high salaries. This is partially due to the fact that the company does not employ a representative fraction of very young employees, who are more likely to work part-time and at lower wages.

3 Empirical Results

3.1 401(k) Enrollment

We first examine the impact of the active-decision regime on enrollment in the 401(k). Figure 1 plots the fraction enrolled in the 401(k) after three months of tenure for employees who were hired in the first seven months of 1997 (the active-decision cohort) and the first seven months of 1998 (the standard-enrollment cohort). We use the third month of tenure because it could take up to three months for enrollments to be processed in the active-decision regime.⁷ The average three-month enrollment rate is 69% for the active-decision cohort, versus 41% for the standard-enrollment cohort, and this difference is statistically significant at the 1% level for every hire month.

Figure 2 plots the fraction of employees who have enrolled in the 401(k) plan against tenure. The participation rate of the active-decision cohort is a sizeable 28 percentage points higher than that of the standard-enrollment cohort at 3 months of tenure. Over time, the participation rate of the active-decision cohort grows more slowly than that of the standard-enrollment cohort, so the enrollment gap decreases with tenure. Nonetheless, the participation rate of the active-decision cohort exceeds that of the standard-enrollment cohort by 17 percentage points at 24 months of tenure, and by 5 percentage points at 42 months. The participation difference between the two cohorts are statistically significant at the 1% level for every tenure level after the first month.

Figures 1 and 2 could be misleading if enrollees under the active-decision regime are

⁷Enrollments were only processed on the first of each month under the active-decision regime. Since employees had 30 days to turn in their form, an employee who was hired late in a month and turned in her form just before the deadline could be enrolled in the third month after her hire.

subsequently more likely to stop contributing to the 401(k) plan. However, attrition rates from the 401(k) plan are indistinguishable under the active-decision regime and the standard-enrollment regime. Indeed, 401(k) participation is a nearly absorbing state under either enrollment regime.⁸

We ascribe the active-decision effect to the fact that active-decision employees had to express their 401(k) participation decision during their first month of employment, rather than being able to delay taking action indefinitely. However, there is another distinction between the active-decision and standard-enrollment regimes as implemented at the company. Under the standard-enrollment regime, employees could enroll in the 401(k) plan at any time. Under the active-decision regime, if employees did not enroll in the plan in their first 30 days at the company, their next enrollment opportunity did not come until January 1 of the following calendar year.⁹ Therefore, in addition to the required affirmative or negative enrollment decision, the active-decision cohort faced a narrower enrollment window than the standard-enrollment cohort. In theory, this limited enrollment window could cause higher initial 401(k) enrollment rates by accelerating the enrollment of employees who would have otherwise enrolled between the third month of their tenure and the following January.

However, the enrollment differences between the cohorts are too large to be explained by a window effect. If only a window effect were operative, enrollment fractions for the two groups should be equal after twelve months of tenure. In fact, the enrollment fraction of the active-decision cohort at three months of tenure is not matched by the standard-enrollment cohort until 30 months of tenure.

3.2 401(k) Contribution Rate

Although active decisions induce earlier 401(k) enrollment, this may come at the cost of more careful and deliberate thinking about how much to save for retirement. We now turn our focus to the impact of active decisions on the 401(k) contribution rate.

Figure 3 plots the relationship between tenure and the average 401(k) contribution rate for the active-decision and standard-enrollment cohorts. The averages include both participants

⁸These calculations are available from the authors.

⁹In fact, the active-decision cohort we analyze (January to June 1997 hires) was able to enroll in the 401(k) plan any time after November 1997, when the company switched to the phone-based daily enrollment system. At hire, however, the active-decision employees were not aware of this impending change and would have believed January 1, 1998 to be their next enrollment opportunity.

(who have a non-zero contribution rate) and non-participants (who have a zero contribution rate). Because our longitudinal data do not start until September 1997, the contribution-rate profile cannot be computed for the entire active-decision cohort until 9 months of tenure.

The active-decision cohort contributes 4.8% of income on average at month 9, and this slowly increases to 5.5% of income by the fourth year of employment. In contrast, the standard-enrollment cohort contributes only 3.6% of income on average at month 9, and it takes more than 33 months for it to match the active-decision cohort's nine-month savings rate. At each tenure level in the graph, the difference between the groups' average contribution rates is statistically significant at the 1% level.

Figure 4 plots the average contribution rate of employees who have a non-zero contribution rate (i.e., 401(k) participants). In contrast to Figure 3, active-decision participants have a lower average contribution rate than standard-enrollment participants until the fourth year of tenure.¹⁰ To gain insight into this pattern, we plot the 25th, 50th, 75th, and 90th percentile contribution rates for the standard-enrollment and active-decision cohorts in Figure 5. Non-participants are assigned a zero contribution rate and are included in these distributions. We see that at each of these points in the distribution, the active-decision cohort's contribution rate matches or exceeds the standard-enrollment cohort's contribution rate at virtually every tenure level. There is nearly no gap between the two cohorts' contribution rates at the 90th percentile, where enrollment in the 401(k) occurs immediately for both groups. As we move down the savings distribution, the difference between the two cohorts tends to increase, and most of this difference is due to active-decision cohort employees signing up for the 401(k) plan earlier in their tenure. Overall, it seems that employees save at roughly the same rate under both regimes once they have enrolled. Therefore, the lower average contribution rate among active-decision participants is not due to active decisions lowering the savings rates of those who would have otherwise contributed more under standard enrollment. Rather, active decisions bring employees with weaker savings motives into the participant pool earlier in their tenure – a selection effect.

Table 3 presents the results of a tobit regression of the two regimes' contribution rates on demographic variables. The contribution rate is censored below at 0% and above at 17% of pay. Both active-decision and standard-enrollment employees are included in the regression,

¹⁰These differences are statistically different at the 1% level through the 29th month of tenure, and at the 5% level through the 30th month of tenure.

regardless of participation status. If the employee was hired under the standard-enrollment regime, the dependent variable is equal to the contribution rate at 30 months after hire. If the employee was hired under the active-decision regime, the dependent variable is equal to an estimate¹¹ of the contribution rate at 3 months after hire. The explanatory variables are a constant, gender, marital status, log of salary, and age dummies. The effect of these variables is allowed to vary between the active-decision and standard-enrollment cohorts. To test the hypothesis that savings rates are more haphazard under active decisions, we also allow the variance of the error term to vary across the two cohorts.¹²

The regression coefficients suggest that in expectation, there is little difference between the savings rate an employee chooses immediately after hire under active decisions and the rate she would have in effect 30 months after hire under standard enrollment. The only variable we can statistically reject having the same effect under both regimes is gender; at our company, women save somewhat less than men under active decisions but not under standard enrollment. The variance of the error term in the regression is significantly smaller for the active-decision cohort than for the standard-enrollment cohort, suggesting that the rush of the active-decision deadline does not cause people to make more haphazard savings-rate decisions.¹³

In sum, active decisions cause employees to immediately choose a savings rate that on average they would take up to 30 months to attain under standard enrollment.

3.3 401(k) Asset Allocation

The effect of active decisions on asset allocation cannot be cleanly inferred from the natural experiment that we study because the menu of investment fund options changed in November 1997, the same time that the company switched from active decisions to the standard-enrollment regime. Prior to the change, employer stock was only available as an investment

¹¹This estimate is constructed by taking the earliest-available contribution rate (which may be as late as 9 months after hire) for the active-decision employee. Naturally, we set that contribution rate to zero if the employee had not enrolled in the plan within 3 months of hire.

¹²To equalize sample selectivity for the two cohorts, we restricted both samples to employees who remain in our data for 30 months. This is why the number of data points in the regression is less than the total number of employees in the two cohorts.

¹³On the other hand, if idiosyncratic variation in the savings rate is primarily due to employees optimally utilizing private information that is not available to the researchers, then a lower error-term variance could be consistent with a less well-considered savings decision.

option for the 14% of participants who made contributions with after-tax money; such after-tax contributions are infrequent because they do not garner matching contributions at this company and are generally less tax-efficient than contributions made with pre-tax money.¹⁴ During the transition to standard enrollment, employer stock was added as an investment option for pre-tax 401(k) contributions. Subsequently, the average allocation to employer stock more than doubled and the average allocation to all other asset classes decreased. It is impossible to determine how much of this increase was caused by the standard-enrollment regime, and how much was caused by the roughly seven-fold increase in the fraction of employees for whom employer stock was a viable investment option.

The impact of active decisions on asset allocation is an important open question, since employees have low levels of financial knowledge about different asset classes (John Hancock, 2002) and many tend to make poor asset-allocation choices (Benartzi and Thaler, 2001; Cronqvist and Thaler, 2004). We discuss later in the paper why active decisions are potentially better suited for contribution-rate choices than for asset-allocation choices.¹⁵

3.4 401(k) Asset Accumulation

We next consider the impact of active decisions on asset accumulation. Unfortunately, any asset-accumulation analysis of cohorts who began savings at different points in time will be confounded by time effects, since asset returns are highly volatile. Moreover, as explained above, the fund menu changed over time, further confounding this analysis. Nonetheless, it is the level of asset accumulation that will ultimately drive retirement timing and consumption levels. Studying asset accumulation also gives us insight into whether increased contribution rates under active decisions are offset by increased 401(k)-loan activity and withdrawals.¹⁶

To measure asset accumulation, we divide 401(k) balances by annual base pay. Our measure of 401(k) balances excludes outstanding principal from 401(k) loans and any balances an employee rolled over from a previous employer.

Figure 6 reports balance-to-pay ratios at the 25th, 50th, 75th, and 90th percentiles of

¹⁴Pre-tax contributions are more tax-efficient unless the contributor has a short investment horizon and expects tax rates to rise sharply in the future.

¹⁵We believe that defaults are usually optimal for asset allocation decisions.

¹⁶The active-decision cohort did not have 401(k) loans available to them at the time they made their initial contribution-rate decision. However, after November 1997, they were able to borrow against their 401(k) balances.

the balance-to-pay distribution for the active-decision and standard-enrollment cohorts. The impact of the market downturn in 2001 appears around the 48th month of tenure for the active-decision cohort and the 36th month of tenure for the standard-enrollment cohort.

It is apparent that the balance-to-pay ratio paths are nearly identical for the two cohorts at both the 75th and 90th percentiles. In contrast, the 25th percentile active-decision employee has a much higher balance-to-pay ratio because participation begins two years earlier in her tenure than it does for the 25th percentile standard-enrollment employee. The 50th percentile active-decision employee has a slightly higher balance-to-pay ratio, but the effects of the 2001 stock market downturn muddy the picture. Overall, it appears that active-decision enrollment only affects asset accumulation in the bottom half of the accumulation distribution. This is consistent with the results of the contribution rate analysis.

4 A Model of 401(k) Enrollment

The empirical analysis in Section 3 shows that requiring an active decision accelerates 401(k) enrollment. But these results do not enable us to evaluate the welfare consequences of active decisions. Active decisions may be harmful if enrolling is a costly action, and there is option value associated with waiting for a more convenient time to enroll. We now present a structural model that provides a framework for thinking about socially optimal enrollment regimes. Each regime is expressed as a default contribution rate in our model. The possible optimal regimes are standard enrollment (a default contribution rate of 0%), automatic enrollment (a positive default contribution rate), or active decision (which is conceptually equivalent to a default so extreme that all employees will choose to opt out immediately).

Our theoretical analysis consists of two parts. The first part is a model of procrastination by the individual employee. There are two key features to this model: a stochastically time-varying transactions cost for opting out of the default, which creates an option value to waiting; and present bias, which leads to procrastination, so that the employee may fail to maximize her expected utility. Although we focus on 401(k) enrollment, this model can be applied to any situation in which an action that generates a positive net utility flow is delayed by procrastination. The key prediction is that procrastination can generate more inefficiency when the utility flow from acting is small than when the utility flow is large, since a large utility-flow benefit will motivate an agent to overcome procrastination and act

immediately.

We then solve the benchmark problem of optimal default-setting by a social planner – for instance, a regulator that is trying to maximize social surplus).¹⁷ We assume that optimal savings rates are heterogeneous across employees and only privately observed. These assumptions are crucial for our welfare results. If the planner knows each employee’s optimum, then the planner can simply set that optimum as the default, sparing the worker the time and effort of making the choice herself. There is a growing body of evidence that planners make better *asset-allocation* choices than workers (Benartzi and Thaler, 2001; Cronqvist and Thaler, 2004). However, survey evidence suggests that workers have idiosyncratic *savings* needs, and that workers understand this individual variation (Choi et al, 2002). Hence, the model that follows is best applied to savings-rate choices, where individuals have superior information about their optimum, and not to asset-allocation choices, where outside experts might have superior information (Benartzi and Thaler, 2001; Cronqvist and Thaler, 2004).

4.1 The Sophisticated Employee’s Problem

We model the employee as an agent who suffers a potential utility loss of $L \geq 0$ at the beginning of each period until she takes some action. In our particular application, the relevant action is opting out of the 401(k) default contribution rate and moving to a (personally) preferred contribution rate. This action entails an immediate transactions cost which is a random variable whose density function $f(c)$ is continuous with support $[\underline{c}, \bar{c}]$ (where $0 < \underline{c} < \bar{c}$).¹⁸

In every period, the agent first experiences a flow loss of L , and then she draws a cost c from the transaction cost distribution. (We assume costs are independent across periods.) She then decides whether to take action. If she takes action, she incurs an immediate cost of c but incurs no further flow losses. If she does not take action, she faces another loss of L at the beginning of the next period, and the game continues. In the 401(k) enrollment context, L represents the net disutility per period due to either oversaving or undersaving at

¹⁷In reality, any institution with default-setting power may have complex motives that extend beyond maximizing social welfare.

¹⁸When $\bar{c} = \infty$, much of the same analysis goes through with technical adjustments. If the distribution of c has a finite mean, the results are very similar, but active decisions (discussed in Section 4.3) must be treated as a separate option rather than an extreme default. If the distribution is sufficiently fat-tailed that the mean of the cost distribution is infinite, then active decisions are no longer a candidate for the optimum, since they impose an infinite average cost.

the default contribution rate (relative to her personal optimal contribution rate). We assume for simplicity that L is constant over time.

The agent has quasi-hyperbolic preferences (Phelps and Pollak 1968, Laibson 1997): she has a long-term discount factor δ and an *additional* short-term discount factor β , where $\beta, \delta \in (0, 1]$. Thus, if her utility in periods $0, 1, 2, \dots$ is u_0, u_1, u_2, \dots , then her intertemporal utility from the point of view of self t is $U_t = u_t + \beta(\delta u_{t+1} + \delta^2 u_{t+2} + \delta^3 u_{t+3} + \dots)$. If $\beta < 1$, this gives rise to dynamic inconsistency. The agent's behavior is modeled as an extensive-form game among the various period- t selves, for all $t \geq 0$. The period- t self's payoff is U_t . We assume for now that the agent is a sophisticated quasi-hyperbolic discounter, meaning that the values of β and δ are common knowledge among the various selves.

In any strategy profile, we can define the payoff of the period- t self if she chooses not to take action.¹⁹ It is then clear that she will take action if and only if she draws a cost c less than this payoff. That is, the strategy in each period is to determine a cutoff cost c^* , and take action if and only if $c < c^*$. We confine our analysis to stationary equilibria, where the cutoff c^* is the same in each period. In the exposition below, we defer most proofs to the appendix.

Proposition 1 *The agent's game has a unique stationary equilibrium:*

- if $L \leq \underline{c}(1 - \delta)/\beta\delta$, then $c^* = \underline{c}$ (the agent never acts);
- if $L \geq \bar{c}/\beta\delta - E(c)$, then $c^* = \bar{c}$ (the agent acts immediately no matter what the cost);
- otherwise, c^* is the unique solution to the equation

$$c^* = \beta\delta (L + E(c | c < c^*))P(c < c^*) + (c^*/\beta)P(c > c^*). \quad (1)$$

In fact, one can show that if β is not too low, this is the *only* equilibrium of the game; there are no non-stationary equilibria. The appendix contains some further discussion.

We are interested not only in the agent's behavior but also in the resulting welfare consequences. We take as our normative welfare measure the expression $u_0 + \delta u_1 + \delta^2 u_2 + \dots$, the stream of losses as discounted with the exponential discount factor δ . Let ϕ denote the

¹⁹We only study Markov strategies—in other words, strategies that depend only on payoff-relevant information. In particular, this excludes equilibria in which the agent's period t strategy depends on the costs drawn in earlier periods.

expectation of this value.²⁰ If the worker's first cost draw is less than c^* , then her total realized loss is c ; otherwise, she incurs an expected loss of $L + \phi$ starting from the next period. So her expected total loss satisfies

$$\phi = E(c | c < c^*)P(c < c^*) + \delta(L + \phi)P(c > c^*),$$

implying that

$$\phi = \frac{E(c | c < c^*)P(c < c^*) + \delta L \cdot P(c > c^*)}{1 - \delta P(c > c^*)}. \quad (2)$$

When $c^* = \underline{c}$, we have $\phi = L\delta/(1 - \delta)$; when $c^* = \bar{c}$, we have $\phi = E(c)$. The intermediate region is of most interest.

We summarize some simple comparative statics on c^* and ϕ . We assume that δ and the distribution of c are given, and consider variation in β and L .

Proposition 2 1. *In the region where $\underline{c} < c^* < \bar{c}$, c^* is strictly increasing in L and β .*

2. *ϕ is weakly decreasing in β .*

3. *If $\beta = 1$, then ϕ is weakly increasing in L . However, if $\beta < 1$, then there exist values of L for which $\phi > E(c)$.*

4. *If $\delta = 1$ and β is sufficiently low, then $\phi \geq E(c)$ for all $L \neq 0$.*

For time-consistent agents ($\beta = 1$), a higher flow loss L always implies a weakly higher expected total loss ϕ . However, agents with $\beta < 1$ will set a lower action threshold c^* than agents with $\beta = 1$, and thus will take longer to act on average. We think of this incremental delay as procrastination. Part 2 of the preceding proposition shows that procrastination leads to lower welfare. In particular, Part 3 shows that there exist some values of L such that the agent's expected losses are larger than the average transactions cost. This implies that an agent facing such an L would actually be better off if L were increased to the point that she would be motivated to take action immediately. Part 4 gives a stronger result: when $\delta = 1$ and time-inconsistency is sufficiently strong, the agent is *always* better off being forced to act immediately if he is not already at his optimum.

²⁰To motivate this long-run perspective, it is enough to assume that regulations established by the planner in period t take effect in period $t + 1$. Then every worker at every point in time will want the planner to set the policy that minimizes ϕ .

For “reasonable” shapes of the transactions cost distribution, the total expected loss ϕ is a single-peaked function of L when $\beta < 1$, rising for low L and then falling, but this is not always true.²¹

To illustrate Proposition 2, Figure 7 graphs the expected total loss ϕ as a function of L for various parameter sets in the case where c is uniformly distributed. When L is close to zero, the expected loss is always increasing in L . When L is sufficiently large, the worker will always immediately opt out of the default at any cost ($c^* = \bar{c}$) and never incur any flow losses. Thus for large L , the total loss is $E(c)$, which is independent of L ; this is why the graphs flatten out on the right.

When $\beta = 1$ (the left graph in Figure 7), ϕ is always weakly increasing as L moves away from zero; time-consistent workers are always weakly better off with a smaller flow loss. But when $\beta < 1$ (the middle graph in Figure 7), there is an intermediate region in which $\phi(L) > E(c)$ (part 3 of Proposition 2). Workers in the “hump” of the loss function would be better off if L were much larger. A larger flow loss is a kick in the pants that causes workers to overcome procrastination.

When L is sufficiently close to (but not equal to) 0, never opting out of the default will be the efficient choice when $\delta < 1$. For these employees, even if they were perfectly time-consistent, the perpetual stream of losses would not be serious enough to warrant incurring the opt-out cost. In this region, ϕ increases linearly with L . The region vanishes as $\delta \rightarrow 1$ or $\underline{c} \rightarrow 0$. In the case $\delta = 1$, if β is sufficiently low (the right graph in Figure 7), then all employees except those at $L = 0$ are made weakly better off by being forced to act immediately; this is part 4 of Proposition 2.

4.2 The Planner’s Optimization Problem

We now describe the problem of a social planner. Suppose that each worker has a private optimal savings rate s . Over the population of workers, s has some distribution with density $g(s)$ and compact support $[\underline{s}, \bar{s}]$. Workers who prefer not to be enrolled in the 401(k) have s close to 0, or $s < 0$ if they would like to dissave. The planner sets a default savings rate

²¹A necessary and sufficient condition for ϕ to have this shape is that $\beta < 1$ and either the equation $c^*f(c^*) = P(c > c^*)/(1 - \beta)$ has exactly one solution in (\underline{c}, \bar{c}) , or else it has no solution and $\delta < 1$. This can be verified using the computation of $d\phi/dc^*$ given in the proof of Lemma 7 in the appendix, and by the observation that when $\delta < 1$, ϕ is always increasing for sufficiently small values of L where the agent never opts out. In particular, the condition is satisfied when c is uniformly distributed and β is not too low.

d . Each worker would like to be at her optimal savings rate s and suffers a utility loss L for each period in which she remains at the default d if $s \neq d$. Workers probabilistically opt out of the default and set their savings rate to their personal optimum s , according to the model described in the previous section.

We assume that L depends only on the difference $\Delta = s - d$. Let $L = l(\Delta)$, and assume

- l is continuous;
- $l(\Delta)$ is decreasing on $(-\infty, 0)$ and increasing on $(0, \infty)$;
- $l(0) = 0$, and $l \rightarrow \infty$ as $\Delta \rightarrow \pm\infty$.

We assume for simplicity that all workers have the same (β, δ) and distribution of transactions costs; only s varies across workers. We can unambiguously write the individual worker's expected total loss as $\phi(l(s - d))$, which we more succinctly call $\Phi(s - d)$. If the default is sufficiently far outside of $[\underline{s}, \bar{s}]$, then all workers are guaranteed to opt out of the default immediately. Such a default is equivalent to an active-decision regime, where workers are required to choose their optimal savings rate immediately.

These additional assumptions on L make the model of the individual employee operational enough that we can check if it fits the data reasonably well. We had assumed that each individual's L is time-invariant. If $L = l(\Delta)$ and the default d is fixed, then the optimal savings rate s must also be fixed. This is consistent with Figure 5 and Table 3, which showed that the contribution rate employees choose when forced to immediately state their preference is similar to the contribution rate they would have eventually chosen under standard enrollment—implying that the optimal savings rate shortly after hire is similar to the optimum at higher tenures.

The individual employee model also predicts that as L increases, the average time to opting out decreases. Figure 8 shows that the data support this prediction. In the standard-enrollment cohort – i.e. the case $d = 0$ – the mean days between hire and enrollment decreases with s (the contribution rate chosen by the worker when they enroll), and hence with $s - d$.

The planner's problem is to choose d to minimize

$$\int_{\underline{s}}^{\bar{s}} g(s)\Phi(s - d) ds. \tag{3}$$

Compactness ensures an optimum exists. Importantly, we are assuming that the planner cannot observe individuals' optimal savings rates and only knows the density function $g(s)$. Otherwise, the planner could simply default each employee into her own optimal savings rate s .

In the simple case where the optimal savings rate s is uniformly distributed, the planner's optimization problem is to minimize the integral of Φ from $\underline{s} - d$ to $\bar{s} - d$ — that is, to pick out the “window” of width $\bar{s} - \underline{s}$ in which the integral of Φ is lowest. The position of the window is determined by d . For more general distributions of s , the problem is similar, but now different parts of the window may be weighted differently.

As with the first part of our model (the employee's decision), we have described the problem of optimal defaults in the context of 401(k) plans, but a similar setup could be applied to any context in which a planner must choose a default. It is also worth mentioning that the default-setting problem interacts with the employee's problem only through the function ϕ , whose relevant qualitative features are summarized in Proposition 2. In particular, we will study how the value of β affects the optimal choice of default, but our results will not depend on the details of how the employee's behavior is modeled. For example, if β were a parameter describing status quo bias (Kahneman, Knetsch, and Thaler, 1994; Samuelson and Zeckhauser, 1988) rather than discounting, similar results would obtain.

4.3 Optimality of Active Decisions

Our analysis can be used to determine the optimality of an active-decision regime. Since the aggregate welfare loss from such a regime is $E(c)$, the question boils down to whether there exists d so that (3) is less than $E(c)$.

There are five parameters in our model: the discounting parameters β, δ ; the loss function l ; and the distributions of transactions costs and optimal savings rates (f and g , respectively). The following two propositions state that regions of the parameter space where active decisions are optimal tend to be regions where time-inconsistency is stronger.

Proposition 3 *If active decisions are optimal at some value of the parameters, then if β is lowered (and everything else remains the same), active decisions are still optimal.*

Proof. The statement that active decisions are optimal means that, for every choice of d ,

$$\int_{\underline{s}}^{\bar{s}} g(s)\Phi(s-d) ds \geq E(c). \quad (4)$$

By Proposition 2, when β decreases, $\Phi(s-d)$ can only increase, so the left side of (4) can only increase, and the inequality stays true. ■

Proposition 4 *Hold $\delta, f,$ and l fixed. If β is sufficiently close to 1, then for any distribution of optimal savings rates, active decisions are never optimal.*

An intuitive companion result would be that whenever β is low enough, active decisions are always optimal. This is true if $\delta = 1$. However, if $\delta < 1$, never opting out may be optimal for agents if the default is already very close to their s . Hence, if the population is highly concentrated near one point, it may be optimal to make that point the default rather than forcing everybody to opt out immediately. The next proposition states these results formally.

Proposition 5 *Fix $f, l,$ and δ .*

1. *Suppose $\delta = 1$. If β is sufficiently low, then active decisions are always optimal.*
2. *Suppose $\delta < 1$. If $\bar{s} - \underline{s}$ is sufficiently small, then the optimal default will always lie between \underline{s} and \bar{s} (so the optimum is not active decisions).*

We will sometimes refer to defaults $d \in [\underline{s}, \bar{s}]$ as *internal* defaults.

In the case where $\delta < 1$, it is difficult to be more precise about whether or not active decisions are optimal; we need some parameter that measures, in the relevant way, how homogeneous the population's optimal savings rates s are. If we impose the structural restriction that s is *uniformly* distributed—an assumption we will maintain in the next section—then we can be more specific:

Proposition 6 *Fix $f, l,$ and $\delta < 1$. If s is uniformly distributed, then there exists a bound w^* with the following property: when β is sufficiently low, active decisions are optimal when $\bar{s} - \underline{s} > w^*$, and an internal default is optimal when $\bar{s} - \underline{s} < w^*$.*

The message of these last two propositions is that when β is low, active decisions will typically be optimal, except when preferences are concentrated.

When active decisions are not optimal, there typically is a range of different optimal default regimes. In order to give an explicit description of these regimes that maps intuitively into 401(k) enrollment mechanisms used in practice, we need to impose more structure on the model. We do this in the next subsection.

4.4 Characterizing Optimal Default Policies

We now assume that transactions costs are uniformly distributed on the interval $[\underline{c}, \bar{c}]$, and optimal savings rates are uniformly distributed on $[\underline{s}, \bar{s}]$. This puts parametric structure on f and g . For the loss function l , we assume $l(\Delta) = \kappa \cdot \Delta^2$, where $\kappa > 0$ is some constant.²² Finally, we assume for the remainder of this section that $\delta = 1$.²³

In this case, we can describe the shape of the expected total loss function Φ . The function is symmetric around $\Delta = 0$ because of the functional form assumption on $l(\Delta)$. Two other properties are stated in the following lemma.

Lemma 7 *Under the assumptions of this section,*

- *If $2 - \bar{c}/\underline{c} < \beta < 1$, there exist $0 < \Delta_m < \bar{\Delta}$ such that $\Phi(\Delta)$ is increasing on $[0, \Delta_m]$, decreasing on $[\Delta_m, \bar{\Delta}]$, and constant at $E(c) = (\underline{c} + \bar{c})/2$ on $[\bar{\Delta}, \infty]$.*
- *If $\beta \leq 2 - \bar{c}/\underline{c}$, then there exists $\bar{\Delta}$ such that $\Phi(\Delta)$ is decreasing on $(0, \bar{\Delta}]$ and constant at $E(c)$ on $[\bar{\Delta}, \infty]$.*

The explicit values are given in the appendix. Figure 9 graphs Φ for three sets of parameter values. (Ignore the shading in the figure for now.) When $\beta = 1$ (the left panel), Φ is (weakly) monotonically increasing in $|\Delta|$; this is immediately given by Proposition 2. For “typical” parameter values when $\beta < 1$, Φ has the two humps shown in the middle and right

²²In firms that match employee 401(k) contributions up to a threshold, the utility-loss function may be kinked at this threshold (which may or may not coincide with the worker’s optimum). In this case, our analytically convenient loss function is unrealistic.

²³When $\delta < 1$, the results are similar, but for certain parameter values there may exist a fourth type of default aside from the three mentioned in Proposition 8. We avoid this case because it makes the algebra significantly more complex without offering new economic insights.

panels. If the range of transactions costs is small and β is low, then Φ is strictly decreasing as Δ moves away from 0, and there is only one hump. (This case is not shown in Figure 9.)

It follows by continuity that there will be at most one value $\Delta_e \in (0, \bar{\Delta})$ such that $\Phi(\Delta_e) = E(c)$. (The value of Δ_e is given in the appendix.) The middle panel of Figure 9 is an example where such a Δ_e value does exist in the inner region of the right hump (with a symmetric $-\Delta_e$ value in the inner region of the left hump). If β is sufficiently low, then $\lim_{\Delta \rightarrow 0} \Phi(\Delta) \geq E(c)$, and no such Δ_e will exist, as in the right panel of Figure 9.

The following proposition shows that when $\beta < 1$, we can classify the planner's solution into one of three cases.

Proposition 8 *If $\beta < 1$, then the optimal default is one of the following three types:*

- the center default $d = (\underline{s} + \bar{s})/2$;
- an offset default, such that $\underline{s} - d = -\Delta_e$ while $\bar{s} - d > \bar{\Delta}$ (or its symmetric equivalent, $\bar{s} - d = \Delta_e$ and $\underline{s} - d < -\bar{\Delta}$);
- active decisions, which correspond to any d with $\underline{s} - d \geq \bar{\Delta}$ or $\bar{s} - d \leq -\bar{\Delta}$.

A detailed proof is in the appendix, but a sketch is as follows. We first show that if $0 < a < b$ and $\Phi(a) = \Phi(b)$, then $|\Phi'(b)| > |\Phi'(a)|$; that is, the outer portion of the humps is steeper than the inner portion. Next, the first-order condition for d is simply that $\Phi(\underline{s} - d) = \Phi(\bar{s} - d)$, leading to a finite number of candidate optima. Finally, using the above fact about the slope of Φ on the humps, we show that any d outside the three classes cited would fail to satisfy the second-order condition.

The possible optimal defaults correspond to the different panels of Figure 9. The area of the shaded regions equals the total population welfare loss from the chosen default, and their width equals $\bar{s} - \underline{s}$. The left panel shows the center default, which sets the default at the mean of the population s distribution. The middle panel shows the offset default, which in this case is to the left of the mean s (there is a symmetric offset default to the right of the mean s). The offset default is placed so that workers with the lowest optimal savings rate, \underline{s} , opt out with some probability less than 1 in the first period, but procrastination causes their expected welfare loss to exactly equal the expected welfare loss if they were forced to opt out with certainty in the first period. The offset default also causes workers with the

highest optimal savings rate, \bar{s} , to opt out with certainty in the first period because the default is far away from their optimum. If $\bar{s} - \underline{s}$ is not large enough for the two extremes of the population to act in this manner, then the offset default is not a candidate for the optimal default. Finally, the right panel of Figure 9 shows an active-decision regime, where the default is set so far outside of the support of s that all individuals opt out of the default immediately and incur expected welfare losses of $E(c)$.

Figure 10 shows how the optimal default depends on two parameters in our model: β (the time-inconsistency factor) and $\bar{s} - \underline{s}$ (the heterogeneity of optimal savings rates). We discuss each region in turn.

First consider the southeast region of Figure 10. In this region, employees have weak dynamic inconsistency problems and relatively homogeneous savings rates. The socially optimal solution here is a center default. This puts all employees in the middle of the graph of the function Φ , where the resulting losses are low.

As β falls, the humps in the graph of Φ grow, and eventually, a center default puts so many workers in the humps that it is no longer optimal. Once β is low enough, the losses from procrastination are large relative to the option value of waiting for a low cost, and employees are better on average if they are forced to opt out of the default immediately. Thus, active decisions are optimal in this region.

When β is high—so that the humps are not too large—and employees are very heterogeneous, the best solution is an offset default, which puts employees into the valley between the humps, one hump, and a plateau, rather than in both humps, the valley, and possibly the plateaus, as would be the case with a center default. Under an offset default, some employees (but not all) are so far from their optimal savings rate that they are compelled to opt out of the default immediately. The offset default is thus a compromise between the active-decision and center solutions. By using an offset default, the planner beneficially moves population mass from one of the humps to a plateau, while still letting those with optimal rates near the new default exploit the option value of waiting.

The following proposition shows that the regions of Figure 10 generically have the shape shown.

Proposition 9 *Fix κ , \underline{c} , and \bar{c} . Then there exist values $0 < \beta_{ac} < \beta_{oc} < 1$, and a function $w : (\beta_{ac}, 1] \rightarrow (0, \infty]$, with the following properties:*

1. *for $\beta \leq \beta_{ac}$, active decisions are always optimal;*

2. for $\beta_{ac} < \beta < \beta_{oc}$, active decisions are optimal when $\bar{s} - \underline{s} > w(\beta)$ and a center default is optimal when $\bar{s} - \underline{s} < w(\beta)$;
3. for $\beta_{oc} < \beta < 1$, an offset default is optimal when $\bar{s} - \underline{s} > w(\beta)$ and a center default is optimal when $\bar{s} - \underline{s} < w(\beta)$;
4. w is increasing on $(\beta_{ac}, \beta_{oc}]$.

The boundaries of the regions in Figure 10 are, in general, difficult to describe explicitly. Notice however that the boundary between the active-decision and offset default region is a vertical line. This is because, for both of these policies, some of the workers are in the plateaus of the function Φ . If the range of s widens slightly, the marginal employees (at the boundary of the support of Δ) will be in the plateau under either regime, so increasing the range of s cannot make one regime more attractive than the other.

The case $\beta = 1$ is an exception. In this case, the function Φ has no humps. When the range of savings rates is low, a center default is optimal. When savings rates are wide enough to cover the whole valley in the graph of Φ , then any default that is sufficiently far inside the interval $[\underline{s}, \bar{s}]$ to take full advantage of the valley will be optimal. More precisely:

Proposition 10 Fix κ , \underline{c} , and \bar{c} , and assume $\beta = 1$. Let $\bar{\Delta}$ be the smallest positive value with $\Phi(\bar{\Delta}) = E(c)$. Then

- if $\bar{s} - \underline{s} \leq 2\bar{\Delta}$, then the center default $d = (\bar{s} + \underline{s})/2$ is the unique optimum;
- otherwise, the set of optimal defaults consists of all $d \in [\underline{s} + \bar{\Delta}, \bar{s} - \bar{\Delta}]$.

We should also mention that although we have focused on β and $\bar{s} - \underline{s}$ as the parameters of interest, we also get information about the effect of κ , the flow loss coefficient, for free. It is easy to check that scaling κ by some positive factor λ has exactly the same effect on the planner's optimization problem as scaling \underline{s} , \bar{s} , and d all by $\sqrt{\lambda}$.

4.5 The Case of Naive Workers

We briefly return to our more general problem, without the distributional assumptions of the Section 4.4, to discuss an extension. Our analysis to this point has assumed that workers are sophisticated and understand their own time-inconsistency. O'Donoghue and Rabin (1999a,

b) have considered naive and partially naive agents. Such agents believe their future selves will act with some short-term discount factor $\widehat{\beta} > \beta$.²⁴ In the case of total naiveté, $\widehat{\beta} = 1$, agents simply plan a strategy for all future periods that maximizes their current self's utility and then best-respond to this strategy in the current period, ignoring the fact that future selves will not want to follow the strategy that was naively forecast for them.

As one would expect, for a fixed β , naive agents do worse than sophisticated agents. The more naive the agent is, the less she expects to lose in the future, and the lower she sets her cutoff cost threshold c^* , which in turn makes her actual losses even higher. The next proposition states this formally.

Proposition 11 *In the employee model, if f , β , δ , and L are held fixed, then c^* is (weakly) decreasing in $\widehat{\beta}$, and ϕ is (weakly) increasing in $\widehat{\beta}$.*

This leads immediately to the following corollary for the social planner:

Corollary 12 *If active decisions are optimal at given parameter values, then they remain optimal when $\widehat{\beta}$ is increased.*

Proof. Exactly the same argument as for Proposition 3. ■

In the extreme case $\delta = 1$, naiveté has some perverse consequences. In particular, sophisticated agents will always set $c^* > \underline{c}$ (unless their flow loss L is zero). However, naive agents with L sufficiently close to 0 will never bother taking action. They expect themselves to take action at some point in the future, and they therefore see no need to act now even if the cost is low.

Proposition 13 *Fix f , β , $\widehat{\beta}$. Assume $\delta = 1$. For all sufficiently small L , the agent will set $c^* = \underline{c}$.*²⁵

Since $\delta = 1$, the total loss ϕ incurred will be infinite. (This phenomenon is also discussed in O'Donoghue and Rabin 1999b.)

This implies that the analysis of our special case (c and s uniformly distributed) is quite different when employees are naive. In this case, an internal default is *never* optimal

²⁴More precisely, for the equilibrium problem to be well-defined, we must assume that for each t , the period- t self believes all future selves are *sophisticated* $(\widehat{\beta}, \delta)$ -discounters. We also continue to assume that the agent expects her future selves to follow a stationary equilibrium.

²⁵Since f is a continuous density function, setting $c^* = \underline{c}$ implies that the agent never acts.

because it causes infinite losses for some agents (unlike active decisions). This result may seem artificial since $\delta = 1$ is an extreme case in our model. But even if we assume $\delta < 1$, so that the losses are finite, similar reasoning will show that internal defaults are still not efficient when δ is close enough to 1. In some cases, the optimal default is neither internal nor active decisions, but a default outside the support of s which does not induce everybody to opt out immediately.

4.6 Delayed Active-Decision Regimes

One more variation we have considered, and which we will discuss only briefly, is that of *delayed* active decisions. In such a regime, the agent knows in period 0 that she must act (and incur the corresponding cost $E(c)$ on average) in period T if she has not already acted before period T . This deadline, T , may be infinite or some finite, nonnegative integer. $T = 0$ is simply the active decision regime that we have already analyzed (the employee must act in the initial period). The employee's problem in the finite- T case is analogous to the infinite- T case.

From the social planner's point of view, a delayed active-decision regime means that the planner sets both a default savings rate d and a deadline T . The active-decision regime at our study company might be described as such a regime, since employees were given 30 days to turn in their forms, rather than being required to turn them in immediately upon joining the company. Intuitively, these regimes have the appeal that they offer some of the option value of waiting without allowing procrastinators to incur excessively large losses.

We do not have a general theory of the planner's solution in situations where T can be positive but finite. However, we do have one result which is surprising in light of the intuition described in the last paragraph:

Theorem 14 *If costs are uniformly distributed, and (β, δ) is homogeneous over the population, then delayed active-decision regimes with finite positive T are never better than $T = \infty$ regimes.*

This holds for any distribution of optimal savings rates. In fact, this holds even if the default d is held fixed; that is, for any distribution of L over the employees (where each individual's L is held fixed), the socially optimal value of T is either ∞ or 0.

The proof is fairly involved, so we omit it here. The key step is that, for each individual, $\phi_T(L)$, the agent’s loss under a T -period regime, converges to $\phi(L)$ as $T \rightarrow \infty$, but the convergence is faster for individuals with $\phi(L) > E(c)$ than for those with $\phi(L) < E(c)$. That is, those individuals who benefit from active decisions (“inefficient” types) converge faster than those who are better off being left to act on their own (“efficient” types). So suppose that $T = 0$ is better than $T = \infty$ for aggregate welfare, so that the inefficient types outweigh the efficient types. Then, for any positive T , the differential rates of convergence imply that the losses to the inefficient types (relative to $T = 0$) exceed the gains to the efficient types, and no such T can be better than $T = 0$. On the other hand, suppose $T = \infty$ is better than $T = 0$ for welfare, so that the efficient types outweigh the inefficient types. Then, for any finite T , the losses to the efficient types (relative to $T = \infty$) outweigh the gains to the inefficient types, and no such T can be better than $T = \infty$.

5 Model Discussion

We now link our theoretical results to actual institutions. We classify *actual* 401(k) enrollment regimes into three types: a standard enrollment 401(k), an automatic enrollment 401(k), and an active decision 401(k).²⁶

Under a standard enrollment 401(k), employees have a default savings rate of zero and are given the option to raise this savings rate. Under an automatic enrollment 401(k), employees have a default savings rate that is strictly positive and are given the option to change that savings rate (including opting out of the plan altogether). Under an active decision 401(k), employees face no default and instead must affirmatively pick a savings rate (which includes the option of saving nothing at all).

In the analysis of our model—see Section 4.4—we found three types of optimal regimes: an offset default, a center default, and active decision. The standard-enrollment 401(k) is an example of an offset default, since a 0% savings rate lies at one end of the optimal savings-rate distribution.²⁷ The automatic enrollment 401(k) may either be an offset default or a

²⁶Contribution escalation (see Thaler and Benartzi, 2004), in which the savings rate increases automatically with some periodicity, is another way in which automatic enrollment is sometimes implemented, and is also included as an opt-in feature in some standard-enrollment 401(k) plans; we do not incorporate the additional complexity of a potentially increasing savings rate here.

²⁷The standard-enrollment default savings rate is on the boundary of the action space, but this location is consistent with the concept of an offset default if savings preferences cross the boundary because some

center default, depending on where the default is set.²⁸ Finally, the active decision 401(k) of the firm studied in the first half of this paper translates directly into the active decision regime derived in our model.

When we calibrate our stylized model we find that active decision regimes are optimal for β values below 0.6 in the sophisticated case and for β values below 0.9 in the naive case. These are plausible values for β ,²⁹ so the calibrations imply that active decisions *might* be socially optimal institutions.

However, such calibrations are only suggestive. To make a compelling quantitative argument we would need a model that was less stylized and more closely tied to deep preference parameters. Hence, a quantitatively meaningful calibration exercise is beyond the scope of the current paper. For such a calibration, we would need a way of estimating the flow losses from sticking to the default. We also need to estimate the density of optimal savings rates, $g(s)$.

Qualitatively, however, we can conclude that three factors robustly strengthen the case for active decision enrollment: greater present bias (lower β), greater naivete, and greater heterogeneity in savings preferences

6 Conclusion

This paper analyzes the active-decision alternative to default-based 401(k) enrollment processes. The active-decision approach forces employees to explicitly choose between the options of enrollment and non-enrollment in the 401(k) plan without advantaging either outcome.

We find that the fraction of employees who enroll in the 401(k) three months after hire is 28 percentage points greater under an active-decision regime than under a standard opt-in enrollment regime. The active-decision regime also raises average saving rates and accumulated 401(k) balances. The distribution of new employees' savings rates under active decisions is similar to the distribution it takes 30 months to achieve under standard enrollment.

households would like to dissave.

²⁸Profit Sharing/401(k) Council of America (2001) reports that three-quarters of companies with automatic enrollment set their default contribution rate at 2% or 3% of pay, which is much lower than the 7% average 401(k) savings rate selected by employees when they make an affirmative choice (Holden and VanDerhei, 2001). However, over the next few years the distribution of default savings rates under automatic enrollment will rise as a result of the passage of the Pension Protection Act.

²⁹See Angeletos et al (2001), and Laibson, Repetto, and Tobacman (2007) for estimates of β .

We also present a general model of procrastination which describes the employee’s choice of a 401(k) contribution rate. Using this framework, we characterize the socially optimal 401(k) regime. The active-decision regime is optimal when workers have relatively heterogeneous savings preferences and a relatively strong tendency to procrastinate.

An active-decision regime has both pros and cons. On the plus side, it is inexpensive to implement, it eliminates costly procrastination, and it pushes heterogeneous workers to choose personally optimal contribution rates. Requiring individuals to make an active decision represents an alternative to the paternalism associated with a planner’s choice of a default. Active-decision interventions are designed principally to force a decision-maker to think about a problem. This is still a type of paternalism, but it does not presuppose a leading answer to the decision problem.³⁰

On the con side, active decisions force workers to engage in a costly decision/implementation process that might be inefficient if a single default would have done a good job for most workers. Likewise, active decisions may force financially unsophisticated workers to make uninformed decisions.

Active-decision interventions will be useful in situations where consumer heterogeneity implies that one choice isn’t ideal for everyone (e.g., the selection of a health plan or automobile insurance³¹) and firms or governments feel uncomfortable implementing employee-specific defaults (e.g., if such employee-specific defaults are viewed as “advice” with fiduciary consequences).³² In contrast, defaults will have a natural role to play in cases where a large degree of homogeneity is appropriate and/or household decision-makers have limited expertise (e.g., portfolio allocation).³³ Future research should explore active-decision experiments in other domains and compare the relative efficacy of active-decision and default-based systems, as well as hybrid systems which integrate their features.³⁴

³⁰We view active decisions as an example of libertarian paternalism (Sunstein and Thaler, 2003).

³¹The active-decision approach to purchasing automobile insurance is widely used. Drivers cannot, in general, register their cars without obtaining insurance. But the government does not specify a default insurance contract for drivers; rather, it requires drivers to obtain their own insurance—to make an active decision. The model in the paper suggests that there is a good justification for this approach: there is likely to be substantial heterogeneity in individual preferences over insurance policy types and companies.

³²An example of an intriguing employee-specific default is a default savings rate that increases with the employee’s age.

³³See Benartzi and Thaler (2003) and Cronqvist and Thaler (2004) for evidence on poor asset allocation choices.

³⁴For instance, the contribution rate in a 401(k) could be an active decision and the asset allocation could be a default.

7 Appendix

7.1 The Sophisticated Employee's Problem

We are concerned mainly with identifying all stationary equilibria of the game, where a fixed cutoff c^* is used in all periods. As in the main text, we let ϕ be the expected total utility loss over all periods, discounted using the exponential factor δ only. Thus ϕ is given by the solution to the Bellman equation, namely expression (2), which we repeat here:

$$\phi = (E(c|c < c^*)P(c < c^*) + \delta L \cdot P(c > c^*)) / (1 - \delta P(c > c^*)).$$

(If $c^* = \underline{c}$ and $\delta = 1$, then we set $\phi = \infty$ for $L > 0$ and $\phi = 0$ for $L = 0$.) The cutoff c^* will be chosen so that that the agent is indifferent between incurring transactions cost c^* and continuing to wait, except that c^* is constrained to lie in the interval $[\underline{c}, \bar{c}]$. That is:

$$c^* = \begin{cases} \underline{c} & \text{if } \beta\delta(L + \phi) < \underline{c} \\ \beta\delta(L + \phi) & \text{if } \underline{c} \leq \beta\delta(L + \phi) \leq \bar{c} \\ \bar{c} & \text{if } \beta\delta(L + \phi) > \bar{c}. \end{cases}$$

Now we can prove the proposition identifying the stationary equilibrium, which we restate here.

Proposition 1 *The agent's game has a unique stationary equilibrium:*

- if $L \leq \underline{c}(1 - \delta)/\beta\delta$, then $c^* = \underline{c}$ (the agent never acts);
- if $L \geq \bar{c}/\beta\delta - E(c)$, then $c^* = \bar{c}$ (the agent acts immediately no matter what the cost);
- otherwise, c^* is the unique solution to the equation

$$c^* = \beta\delta(L + E(c|c < c^*)P(c < c^*) + (c^*/\beta)P(c > c^*)). \quad (5)$$

Proof. The cases of the proposition do not overlap, since $\underline{c}(1 - \delta)/\beta\delta < \bar{c}/\beta\delta - E(c)$. (To see this, note that $\underline{c}(1 - \delta)/\beta\delta < E(c)(1 - \delta)/\beta\delta \leq E(c)/\beta\delta - E(c) < \bar{c}/\beta\delta - E(c)$.)

When is $c^* = \underline{c}$ an equilibrium? This cutoff produces losses of $\phi = \delta L/(1 - \delta)$. This is an equilibrium if and only if this value of ϕ gives $\beta\delta(L + \phi) \leq \underline{c}$, which is equivalent to $L \leq \underline{c}(1 - \delta)/\beta\delta$.

Likewise, a cutoff of $c^* = \bar{c}$ yields $\phi = E(c)$. This cutoff is an equilibrium if and only if $\beta\delta(L + \phi) \geq \bar{c}$, which is equivalent to $L \geq \bar{c}/\beta\delta - E(c)$.

Finally, any given value of c^* in the middle is an equilibrium if and only if $c^* = \beta\delta(L + \phi)$, where ϕ is given by (2). This equation expands to $c^* = \beta\delta(L + E(c|c < c^*)P(c < c^*)) / (1 - \delta P(c > c^*))$. Clearing the denominator and rearranging gives (5).

For each $c^* \in [\underline{c}, \bar{c}]$, let

$$g(c^*) = \beta\delta(L + E(c|c < c^*)P(c < c^*) + (c^*/\beta)P(c > c^*)).$$

To complete the existence-uniqueness proof, we need to check that equation (5), which states that $c^* = g(c^*)$, has exactly one solution on (\underline{c}, \bar{c}) if $\underline{c}(1 - \delta)/\beta\delta < L < \bar{c}/\beta\delta - E(c)$, and none otherwise.

The function g is differentiable:

$$g'(c^*) = \beta\delta c^* f(c^*)(1 - 1/\beta) + \delta P(c > c^*).$$

In particular, $g'(c^*) \leq \delta P(c > c^*) \leq 1$, with equality possible only if $c^* = \underline{c}$ and $\delta = 1$. This implies that $g(c^*) = c^*$ has one root in (\underline{c}, \bar{c}) if $g(\underline{c}) > \underline{c}$ and $g(\bar{c}) < \bar{c}$, and no root otherwise. Since $g(\underline{c}) = \beta\delta L + \delta\underline{c}$ and $g(\bar{c}) = \beta\delta(L + E(c))$, the result follows. ■

What happens if we no longer limit ourselves to stationary equilibria? There is now a separate cutoff c_t^* in each period t , and a separate total loss ϕ_t obtained by δ -discounting all subsequent losses and viewing period t as the present. The value function then becomes

$$\phi_t = E(c|c < c_t^*)P(c < c_t^*) + \delta(L + \phi_{t+1})P(c > c_t^*),$$

and the indifference condition is $c_t^* = \beta\delta(L + \phi_{t+1})$ (with the same truncation as before at \underline{c} and \bar{c}). Combining these, we get $c_t^* = g(c_{t+1}^*)$ (truncated to $[\underline{c}, \bar{c}]$), where g is the same function as above. (There is some extra subtlety when $c_{t+1}^* = \underline{c}$, which we omit here.)

If $|g'(c^*)| < 1$ for all c^* , except possibly at the endpoints \underline{c} and \bar{c} , then contraction mapping theory tells us that the intersection of the images of the iterates g, g^2, g^3, \dots consists of a single point. All the c_t^* lie in all of these images, so they are all equal to this point. Hence, if g' satisfies this condition, the only possible equilibrium is the stationary one.

One can check that if

$$\beta > 1 - \min_{c^* \in [\underline{c}, \bar{c}]} \frac{\delta P(c > c^*) + 1}{\delta c^* f(c^*)},$$

then the condition on g' is met, and so the stationary equilibrium is the only equilibrium. On the other hand, if β is too low, then there may be cyclical equilibria. The no-cyclical-equilibria condition may or may not be met for “most” values of β , depending on the distribution of c . For example, if c is uniformly distributed, the condition reduces to $\beta > 1 - 1/[\delta(1 - \underline{c}/\bar{c})]$, which holds for typical β if the range of transactions costs is wide and δ is high.

We derive an important corollary to Proposition 1:

Corollary 15 *The cutoff c^* of the stationary equilibrium is a continuous function of β, δ , and L , and it is differentiable except at the boundaries $L = \underline{c}(1 - \delta)/\beta\delta$ and $L = \bar{c}/\beta\delta - E(c)$. The same holds for ϕ , except that it is discontinuous when $\delta = 1$ and $L = 0$.*

Proof. Continuity of c^* at the boundary points is straightforward to check. Differentiability of c^* away from the boundary points is easily checked by writing out $c^* = g(c^*)$ and differentiating implicitly with respect to each of the parameters (on which the value of g depends). Continuity and differentiability of ϕ then follow from equation (2), except that the denominator becomes zero when $\delta = 1$ and $c^* = \underline{c}$ (which occurs if and only if $L = 0$), so we cannot infer anything there. Indeed, if $\delta = 1$, we have a discontinuity at $L = 0$: we have set $\phi = 0$ for $L = 0$, but for any nonzero L , $c^* > \underline{c}$, so the agent will eventually act and incur expected costs $\geq \underline{c}$. ■

Next, we prove our comparative statics on c^* and ϕ .

Proposition 2

1. *On the region where $\underline{c} < c^* < \bar{c}$, c^* is strictly increasing in L and β .*
2. *ϕ is weakly decreasing in β .*
3. *If $\beta = 1$, then ϕ is weakly increasing in L . However, if $\beta < 1$, then there exist values of L for which $\phi > E(c)$.*
4. *If $\delta = 1$ and β is sufficiently low, then $\phi \geq E(c)$ for all $L \neq 0$.*

Proof. For part 1, use $c^* = g(c^*)$. If we hold c^* constant but increase either L or β , then $g(c^*)$ increases. Since $g(c^*) - c^*$ is a decreasing function of c^* , the new value c^{**} such that $g(c^{**}) = c^{**}$ must be greater than c^* .

For part 2, first take equation (5) and differentiate implicitly with respect to β . This gives us

$$\frac{\partial c^*}{\partial \beta} = \frac{\delta(L + E(c|c < c^*)P(c < c^*))}{1 + (1 - \beta)\delta c^* f(c^*) - \delta P(c > c^*)} = \frac{c^*(1 - \delta P(c > c^*))/\beta}{1 - \delta P(c > c^*) + (1 - \beta)\delta c^* f(c^*)} \leq c^*/\beta,$$

where the second equality comes from rearranging (5) and substituting into the numerator.

Now rearrange the indifference condition for c^* to get $\phi = c^*/\beta\delta - L$, and differentiate with respect to β :

$$\frac{\partial \phi}{\partial \beta} = \frac{\partial c^*/\partial \beta}{\beta\delta} - \frac{c^*}{\beta^2\delta} \leq 0.$$

Note that equality holds if and only if $f(c^*) = 0$. If f is zero on an interval within the support of c , then ϕ will not be strictly decreasing in β .

For part 3, note that $\partial\phi/\partial L = (\partial c^*/\partial L)/\beta\delta - 1$. If $\beta = 1$, then we can differentiate (5) implicitly with respect to L to get $\partial c^*/\partial L = \delta/(1 - \delta P(c > c^*)) > 0$. Therefore, $\partial\phi/\partial L > 0$.

However, if $\beta < 1$, then consider $L = \bar{c}/\delta - E(c)$. For this L , at $\beta = 1$ we have $\phi = E(c)$. As β falls, by part 2, ϕ will increase (and at least initially will strictly increase, since $f(c^*) > 0$ near \bar{c}).

Finally, for part 4, pick $\beta < \min_{c^*}(c^*/E(c|c > c^*))$. Since $\delta = 1$, $c^* > \underline{c}$ if $L \neq 0$ because never acting leads to an infinite present value of losses. If $c^* = \bar{c}$ then $\phi = E(c)$. Otherwise, the indifference condition gives

$$c^* = \beta(L + \phi) \leq c^*(L + \phi)/E(c|c > c^*),$$

or $E(c|c > c^*) \leq L + \phi$; hence

$$\begin{aligned} \phi &= E(c|c < c^*)P(c < c^*) + (L + \phi)P(c > c^*) \\ &\geq E(c|c < c^*)P(c < c^*) + E(c|c > c^*)P(c > c^*) = E(c). \end{aligned}$$

(Incidentally, the sufficient condition $\beta < \min_{c^*}(c^*/E(c|c > c^*))$ is interesting. It implies that if the range $\bar{c} - \underline{c}$ of possible transactions costs is small compared to the overall magnitude

of the costs, then even very slightly present-biased agents will do worse on their own than under a regime where they are forced to act immediately.) ■

We will use the following couple of facts later:

Corollary 16 *If $L > \bar{c}/\delta - E(c)$, then $\phi(L) \geq E(c)$.*

Proof. Exactly as for part 3 of Proposition 2. ■

Proposition 17 *As $\beta \rightarrow 0$ (and all other parameters stay constant), $c^* \rightarrow \underline{c}$. As $L \rightarrow 0$, $c^* \rightarrow \underline{c}$ also.*

Proof. Both statements are obvious if $\delta < 1$, since then $L < \underline{c}(1 - \delta)/\beta\delta$ if β or L becomes low enough. So assume $\delta = 1$. For the first statement, rewrite (5) as

$$c^* = g(c^*) = \beta(L + E(c|c < c^*)P(c < c^*)) + c^*P(c > c^*)$$

When c^* is close to \underline{c} , this equation can be uniquely solved for β with $0 < \beta < 1$, so c^* is the cutoff used for this β . As $c^* \rightarrow \underline{c}$, $\beta \rightarrow 0$. For the second statement, rewrite (5) again as

$$(c^* - \beta E(c|c < c^*))P(c < c^*) = \beta L$$

As $L \rightarrow 0$, we must have either $c^* - \beta E(c|c < c^*) \rightarrow 0$ or $P(c < c^*) \rightarrow 0$. Either of these implies $c^* \rightarrow \underline{c}$. ■

7.2 Optimality of Active Decisions

In this section we prove three propositions describing whether or not active decisions are optimal in limiting cases.

Proposition 4 *Hold δ, f , and l fixed. If β is sufficiently close to 1, then for any distribution of optimal savings rates, active decisions are never optimal.*

Proof. First, let $\underline{\Delta}, \bar{\Delta}$ be points in the “flat” parts of the graph of $\Phi(\Delta)$, so that $\Phi(\Delta) = E(c)$ whenever $\Delta < \underline{\Delta}$ or $\Delta > \bar{\Delta}$. When $\beta = 1$, we know

$$\int_{\underline{\Delta}}^{\bar{\Delta}} (\Phi(\Delta) - E(c)) d\Delta < 0,$$

since $\Phi(\Delta) = \phi(l(\Delta))$ is always $\leq E(c)$, with strict inequality when Δ is near zero. Using Corollary 15, this integral is continuous in β . So for all sufficiently large β , the integral remains less than 0. Focus attention on these β , and fix $\underline{\Delta}, \bar{\Delta}$ sufficiently far out to work for all such β .

Take any such β , and consider an arbitrary distribution of optimal savings rates with density function g . Consider the integral

$$\int_{\underline{s}-\bar{\Delta}}^{\bar{s}-\underline{\Delta}} \left(\int_{\underline{s}}^{\bar{s}} g(s)[\Phi(s-d) - E(c)] ds \right) dd. \quad (6)$$

By using the fact that the integrand is 0 below $\underline{\Delta}$ and above $\bar{\Delta}$, and changing variables from (d, s) to (Δ, s) , we can rewrite the integral as

$$\int_{\underline{\Delta}}^{\bar{\Delta}} \int_{\underline{s}}^{\bar{s}} g(s)[\Phi(\Delta) - E(c)] ds d\Delta = \int_{\underline{s}}^{\bar{s}} g(s) \left(\int_{\underline{\Delta}}^{\bar{\Delta}} [\Phi(\Delta) - E(c)] d\Delta \right) ds.$$

By choice of β , the value of the inner integral is negative.

Therefore, the value of the integral (6) is negative. So there must exist some value of d for which the inner integral, $\int_{\underline{s}}^{\bar{s}} g(s)[\Phi(s-d) - E(c)] ds$, is negative, which means precisely that this value of d is better than active decisions. ■

Proposition 5 Fix f, l , and δ .

1. Suppose $\delta = 1$. If β is sufficiently low, then active decisions are always optimal.
2. Suppose $\delta < 1$. If $\bar{s} - \underline{s}$ is sufficiently small, then the optimal default will always lie between \underline{s} and \bar{s} (so the optimum is not active decisions).

Proof. The first item is immediate from part 4 of Proposition 2: that proposition says active decision is best for each individual employee, so it is also socially optimal.

The second part essentially follows from the fact that the dip to $\Phi(\Delta) = 0$ at $\Delta = 0$ is the lowest valley in the graph of $\Phi(\Delta)$. We state this more precisely as follows. For $L > 0$, we easily see that $\phi(L) > 0$. Continuity of ϕ means that it has a minimum on the interval $[\underline{c}(1-\delta)/\beta\delta, \bar{c}/\beta\delta - E(c)]$, which necessarily is positive; call it $\underline{\phi}$. Let $\underline{L} = \min\{\underline{\phi}(1-\delta)/\delta, \underline{c}(1-\delta)/\beta\delta\}$. The first quantity is the L that would generate losses $\underline{\phi}$ if the agent never opted out. The second quantity is the L above which the agent opts out with some positive probability.

\underline{L} is in the region where the agent never opts out; ϕ is strictly increasing in L in this region. Hence, for each $L < \underline{L}$, and for all $L' > L$, we have $\phi(L') > \phi(L)$. (If L' starts at L and increases, initially it lies in the region where $c^* = \underline{c}$ and ϕ is increasing; then it is in the region where the agent opts out with positive probability, but ϕ has a minimum of $\underline{\phi}$ in this region, which is greater than $\phi(L)$).

Now, the assumptions on l imply we can choose Δ^* such that $l(\Delta) < \underline{L}$ whenever $|\Delta| < \Delta^*$. We claim that if $\bar{s} - \underline{s} < \Delta^*$, then the optimal default lies inside $[\underline{s}, \bar{s}]$.

Consider any default $d < \underline{s}$. For each s , we have $l(s - \underline{s}) < \underline{L}$ and $l(s - d) > l(s - \underline{s})$. Applying ϕ , we get $\Phi(s - d) > \Phi(s - \underline{s})$. So the default d is strictly worse than a default of \underline{s} for any given s in our range; hence it is socially worse than a default of \underline{s} . Likewise, any default $d > \bar{s}$ is strictly worse than a default of \bar{s} . This completes the proof. ■

Proposition 6 *Fix f, l , and $\delta < 1$. If s is uniformly distributed, then there exists a bound w^* with the following property: when β is sufficiently low, active decisions are optimal when $\bar{s} - \underline{s} > w^*$, and an internal default is optimal when $\bar{s} - \underline{s} < w^*$.*

Proof. Define $\Lambda(\Delta) = l(\Delta)\delta/(1 - \delta)$. The function Λ is the “limit” of Φ as $\beta \rightarrow 0$. (Indeed, for any fixed Δ , $\Phi(\Delta) = \Lambda(\Delta)$ when β is close enough to 0.) Consider all w with the following property: there exists d such that $\int_0^w \Lambda(s - d) ds \leq wE(c)$. (That is, active decisions are not optimal when $\bar{s} - \underline{s} = w$ in the “ $\beta = 0$ ” case.) Let w^* be the supremum of all such w . Clearly w^* is finite. For example, choose $\tilde{\Delta}$ such that $\Lambda(\Delta) > 2E(c)$ when $|\Delta| > \tilde{\Delta}$; then the average of Λ on any interval of width greater than $4\tilde{\Delta}$ is greater than $E(c)$ (since $\Lambda > 2E(c)$ on at least half the interval), showing that $w^* < 4\tilde{\Delta}$. On the other hand, the fact that $\Lambda(\Delta) \rightarrow 0$ as $\Delta \rightarrow 0$ implies that $w^* > 0$.

For each $w < w^*$, let $d(w)$ be a value that minimizes $\int_0^w \Lambda(s - d) ds$. Continuity and compactness arguments show that we can define $d^* = d(w^*)$ so that $\int_0^{w^*} \Lambda(s - d^*) ds \leq w^*E(c)$. Thus the average value of Λ on $[-d^*, w^* - d^*]$ is $\leq E(c)$. In fact, it must equal $E(c)$, or else we could widen the interval slightly, violating the definition of w^* . Similarly, $\Lambda \geq E(c)$ at each endpoint of the interval, or else we could widen the interval and have the average still be $< E(c)$. It is clear from the shape of Λ that for any $w < w^*$, we can push the ends of our interval inward to get an interval of width w where the average value of Λ is $< E(c)$.

Now let β be any value low enough that

- $l(\Delta) < \underline{c}(1 - \delta)/\beta\delta$ when $|\Delta| < w^*$;

- $\bar{c}/\delta - E(c) < \underline{c}(1 - \delta)/\beta\delta$.

Then Φ agrees with Λ on $[-w^*, w^*]$, and has value $\geq E(c)$ outside this interval. (This latter follows from the fact that ϕ is increasing on $[0, \underline{c}(1 - \delta)/\beta\delta]$, and remains $\geq E(c)$ for higher L by Corollary 16.)

So if $w > w^*$, there is no interval of width w on which the average of Φ is less than $E(c)$. (If there were such an interval, trim away the part of the interval outside $[-w^*, w^*]$ and expand the interval if necessary so that it includes 0. The average of Φ in the expansion is less than its average in the trimmed region. This gives us a subinterval of $[-w^*, w^*]$ having width $\geq w^*$ on which the average of Φ is less than $E(c)$, which is impossible.)

And if $w < w^*$, then we have seen that there is a subinterval of $[-d^*, w^* - d^*] \subset [-w^*, w^*]$ on which Λ has average less than $E(c)$. Since Φ agrees with Λ here, the result follows. ■

7.3 Characterizing Optimal Default Policies

As stated in the main text, we now study the special case in which c and s are both uniformly distributed, $l(\Delta) = \kappa\Delta^2$, and $\delta = 1$. We first need a careful characterization of the shape of the function $\Phi(\Delta)$.

Lemma 7 *With the assumptions of the previous paragraph,*

- *If $2 - \bar{c}/\underline{c} < \beta < 1$, there exist $0 < \Delta_m < \bar{\Delta}$ such that $\Phi(\Delta)$ is increasing on $[0, \Delta_m]$, decreasing on $[\Delta_m, \bar{\Delta}]$, and constant at $E(c) = (\underline{c} + \bar{c})/2$ on $[\bar{\Delta}, \infty]$.*
- *If $\beta \leq 2 - \bar{c}/\underline{c}$, then there exists $\bar{\Delta}$ such that $\Phi(\Delta)$ is decreasing on $(0, \bar{\Delta}]$ and constant at $E(c)$ on $[\bar{\Delta}, \infty]$.*

Proof. It suffices to show the analogous statement for $\phi(L)$: if $2 - \bar{c}/\underline{c} < \beta < 1$, there exist $0 < L_m < \bar{L}$ such that $\phi(L)$ is increasing on $[0, L_m]$, decreasing on $[L_m, \bar{L}]$, and constant at $E(c)$ on $[\bar{L}, \infty]$; if $\beta < 2 - \bar{c}/\underline{c}$ then there exists \bar{L} such that $\phi(L)$ is decreasing on $(0, \bar{L}]$ and constant at $E(c)$ on $[\bar{L}, \infty]$. Then $\Delta_m = \sqrt{L_m/\kappa}$ and $\bar{\Delta} = \sqrt{\bar{L}/\kappa}$. Clearly we should set $\bar{L} = \bar{c}/\beta - E(c)$, since ϕ is constant above this \bar{L} .

We wish to compute $d\phi/dL$ for $L < \bar{L}$. To do this, first rewrite equation (2) using the indifference condition (5) for c^* , getting

$$\phi = E(c|c < c^*)P(c < c^*) + (c^*/\beta)P(c > c^*). \quad (7)$$

Incidentally, (7) holds when $\delta < 1$ as well. We can treat this as an expression for ϕ in terms of c^* only. The derivative is

$$\frac{d\phi}{dc^*} = c^* f(c^*) (1 - 1/\beta) + P(c > c^*)/\beta = \frac{c^*(1 - 2/\beta) + \bar{c}/\beta}{\bar{c} - \underline{c}}.$$

This is decreasing in c^* , zero at $c^* = \bar{c}/(2 - \beta) < \bar{c}$ (hence negative at $c^* = \bar{c}$), and positive at $c^* = \underline{c}$ if and only if $\underline{c} < \bar{c}/(2 - \beta)$, which is equivalent to $\beta > 2 - \bar{c}/\underline{c}$.

Now

$$\frac{d\phi}{dL} = \frac{d\phi}{dc^*} \cdot \frac{dc^*}{dL},$$

and c^* is increasing in L (for $L < \bar{L}$) by Proposition 2. So $d\phi/dL$ has the same sign as $d\phi/dc^*$, which was computed in the previous paragraph. It follows that if $\beta > 2 - \bar{c}/\underline{c}$ then ϕ is increasing in L up until the (unique) value for which $c^* = \bar{c}/(2 - \beta)$ (this value exists by Corollary 15 and Proposition 17), and decreasing for higher values of L ; if $\beta \leq 2 - \bar{c}/\underline{c}$ then ϕ is always decreasing in L . ■

From this proof we can directly calculate the values of $\bar{\Delta}$ and Δ_m . First, from \bar{L} , we have

$$\bar{\Delta} = \sqrt{\frac{\bar{c}/\beta - (\bar{c} + \underline{c})/2}{\kappa}}.$$

Next, L_m is the value for which $c^* = \bar{c}/(2 - \beta)$ (if it exists). We can invert equation (5) to get

$$L = (c^*/\beta - E(c|c < c^*))P(c < c^*). \quad (8)$$

Plugging in $\bar{c}/(2 - \beta)$ for c^* , and using $E(c|c < c^*) = (\underline{c} + c^*)/2$ and $P(c < c^*) = (c^* - \underline{c})/(\bar{c} - \underline{c})$ for the uniform distribution, we get

$$L_m = \frac{(\bar{c} - \beta\underline{c})(\bar{c} - (2 - \beta)\underline{c})}{2\beta(2 - \beta)(\bar{c} - \underline{c})} \quad \Rightarrow \quad \Delta_m = \sqrt{\frac{(\bar{c} - \beta\underline{c})(\bar{c} - (2 - \beta)\underline{c})}{2\beta(2 - \beta)(\bar{c} - \underline{c})\kappa}}.$$

We can also calculate the value of Δ_e , the value below $\bar{\Delta}$ for which $\Phi(\Delta_e) = E(c)$, if it exists. We do this by finding L_e such that $\phi(L_e) = E(c)$. Note that at this L_e , (7) implies

$c^*/\beta = E(c|c > c^*)$. Using $E(c|c > c^*) = (\bar{c} + c^*)/2$ gives $c^* = \beta\bar{c}/(2 - \beta)$. Then (8) gives

$$L_e = \frac{\beta\bar{c} - (2 - \beta)\underline{c}}{2(2 - \beta)} \Rightarrow \Delta_e = \sqrt{\frac{\beta\bar{c} - (2 - \beta)\underline{c}}{2(2 - \beta)\kappa}}.$$

Now, as mentioned in the main text, we need to prove that when $\beta > 2 - \bar{c}/\underline{c}$, so that the function Φ has humps, the outer portion of the humps is steeper than the inner portion.

Lemma 18 For $0 < \Delta < \bar{\Delta}$, $d^3\Phi/d\Delta^3 < 0$.

Proof. Write out the condition (5) for c^* explicitly, using the uniform distribution. This condition is a quadratic equation for c^* , whose constant term is a linear function of L and whose other coefficients are independent of L . So c^* can be written in the form $\sqrt{A + BL} + C$, where A , B , and C are constants.³⁵ Next, $\phi = c^*/\beta - L$ (the indifference condition for c^* rearranged) can be written in the form $(\sqrt{A + BL} + C)/\beta - L$, where $A + BL \geq 0$ on the relevant range.

Finally, using $L = \kappa\Delta^2$, we see we can write $\Phi(\Delta)$ in the form $(\sqrt{A + B\kappa\Delta^2} + C)/\beta - \kappa\Delta^2$. By taking $\Delta \rightarrow 0$, we see $A \geq 0$. It is straightforward to check that the third derivative of this function is

$$\frac{-3AB^2\kappa^2\Delta}{\beta(A + B\kappa\Delta^2)^{5/2}} < 0.$$

(We can check that A is strictly greater than zero because differentiating (5) shows that $dc^*/dL \not\rightarrow \infty$ as $L \rightarrow 0$, whereas $A = 0$ implies $dc^*/dL \rightarrow \infty$ as $L \rightarrow 0$.) ■

Proposition 19 If $0 < \Delta_1 < \Delta_2 < \bar{\Delta}$ and $\Phi(\Delta_1) = \Phi(\Delta_2)$, then $\Phi'(\Delta_1) + \Phi'(\Delta_2) < 0$. (Note that Δ_1 is on the inside of the hump and Δ_2 is on the outside.)

Proof. Let $\Psi(\Delta) = d\Phi/d\Delta$. By Lemma 18, Ψ is concave. Jensen's inequality implies that the average value of Ψ on the interval $[\Delta_1, \Delta_2]$ is greater than the average of its value at the two endpoints. But $\int_{\Delta_1}^{\Delta_2} \Psi(\Delta) d\Delta = \Phi(\Delta_2) - \Phi(\Delta_1) = 0$, so the average value of Ψ on the interval is zero. Hence, the average of the values of Ψ at Δ_1 and Δ_2 must be negative. ■

Now we are finally prepared to prove the classification of possible optimal defaults.

Proposition 8 If $\beta < 1$, then the optimal default is one of the following three types:

³⁵The explicit expression is $c^* = (\underline{c} + \sqrt{(1 - \beta)^2\underline{c}^2 + 2\beta(2 - \beta)(\bar{c} - \underline{c})L})/(2 - \beta)$. One can check that the lower root of the quadratic is less than \underline{c} .

- the center default $d = (\underline{s} + \bar{s})/2$;
- an offset default, such that $\underline{s} - d = -\Delta_e$ while $\bar{s} - d > \bar{\Delta}$ (or its symmetric equivalent, $\bar{s} - d = \Delta_e$ and $\underline{s} - d < -\bar{\Delta}$);
- active decisions, which correspond to any d with $\underline{s} - d \geq \bar{\Delta}$ or $\bar{s} - d \leq -\bar{\Delta}$.

Proof. For the purposes of this proof, we redefine $\Phi(0)$ to equal $\lim_{\Delta \rightarrow 0} \Phi(\Delta)$. This makes Φ continuous everywhere and (by Corollary 15) differentiable except at $\Delta = 0, \pm\bar{\Delta}$.

The optimal default is the d that minimizes the average value of Φ on $[\underline{s} - d, \bar{s} - d]$, i.e. that minimizes $\int_{\underline{s}}^{\bar{s}} \Phi(s - d) ds$. As observed in the main text, a minimum always exists. By differentiating with respect to d , we get the first-order condition

$$\Phi(\underline{s} - d) - \Phi(\bar{s} - d) = 0$$

that any optimum must satisfy. The second-order condition is

$$-\Phi'(\underline{s} - d) + \Phi'(\bar{s} - d) \geq 0,$$

which must hold if both derivatives are defined.

Consider now the common value $\Phi(\bar{s} - d) = \Phi(\underline{s} - d)$ that emerges from the first-order condition. We use Lemma 7 to draw conclusions about where $\bar{s} - d$ and $\underline{s} - d$ are located for a given common value:

- If the common value is $< E(c)$, then there are only two points $\pm\Delta$ where Φ takes on this value. So $\bar{s} - d = -(\underline{s} - d)$, and we have a center default.
- If the common value is $> E(c)$, then we may (if $\beta < 2 - \bar{c}/\underline{c}$) again have just two points where Φ takes on this value. At most, there are four such points, of the form $\pm\Delta_1, \pm\Delta_2$, where $0 < \Delta_1 < \Delta_m < \Delta_2 < \bar{\Delta}$. If $\underline{s} - d, \bar{s} - d$ are equal to $\pm\Delta_1$ then we have a center default. Otherwise, assume $\bar{s} - d = \Delta_2$. (By symmetry, the argument is equivalent when $\underline{s} - d = -\Delta_2$.)

If $\underline{s} - d = -\Delta_2$ we have a center default. If $\underline{s} - d = -\Delta_1$, then

$$-\Phi'(\underline{s} - d) + \Phi'(\bar{s} - d) = \Phi'(\Delta_1) + \Phi'(\Delta_2) < 0$$

by Proposition 19. This violates the second-order condition. And if $\underline{s} - d = \Delta_1$ then the entire interval $[\underline{s} - d, \bar{s} - d]$ lies within the hump, so that $\Phi(s) > E(c)$ for each s , and this default is strictly worse than active decision.

- If the common value is equal to $E(c)$, then the possible values for the two endpoints are $\pm\Delta_e$ and $\pm\Delta$ for any $\Delta \geq \bar{\Delta}$. If the endpoints are $\pm\Delta_e$ then we have a center default. Otherwise, without loss of generality, assume $\bar{s} - d \geq \bar{\Delta}$.

If $\underline{s} - d \leq -\bar{\Delta}$ then the interval $[\underline{s} - d, \bar{s} - d]$ contains both humps and the valley between them. We can increase d until the upper endpoint $\bar{s} - d$ hits Δ_e , thus eliminating one hump and replacing it with a plateau. This changes the integral of Φ by $\int_{\Delta_e}^{\bar{\Delta}} (E(c) - \Phi(\Delta)) d\Delta < 0$. Thus, total social loss is decreased, so the original d was not optimal.

If $\underline{s} - d = -\Delta_e$ then we have an offset default. And if $\underline{s} - d = \Delta_e$ then the interval contains one hump and a plateau, which is again strictly inferior to active decisions.

■

Now we have shown that the optimal default types shown in Figure 8 are indeed the only types that can occur, and we are ready to describe the regions where each of them occurs.

Proposition 9 *Fix κ , \underline{c} , and \bar{c} . Then there exist values $0 < \beta_{ac} < \beta_{oc} < 1$, and a function $w : (\beta_{ac}, 1] \rightarrow (0, \infty]$, with the following properties:*

1. for $\beta \leq \beta_{ac}$, active decisions are always optimal;
2. for $\beta_{ac} < \beta < \beta_{oc}$, active decisions are optimal when $\bar{s} - \underline{s} > w(\beta)$ and a center default is optimal when $\bar{s} - \underline{s} < w(\beta)$;
3. for $\beta_{oc} < \beta < 1$, an offset default is optimal when $\bar{s} - \underline{s} > w(\beta)$ and a center default is optimal when $\bar{s} - \underline{s} < w(\beta)$;
4. w is increasing on $(\beta_{ac}, \beta_{oc}]$.

Proof. First, fix any $\beta < 1$. Define $w(\beta)$ to be the supremum of values $\bar{s} - \underline{s}$ for which a center default is optimal, or 0 if a center default is never optimal. This supremum is finite; for example, if $\bar{s} - \underline{s} > 2\bar{\Delta}$, then a center default implies that the interval $[\underline{s} - d, \bar{s} - d]$ contains both humps of the function Φ , and this is strictly worse than the offset default, which contains only one of the humps.

Now suppose a center default is optimal for some width $\bar{s} - \underline{s}$. We claim it is also optimal for all narrower widths. First we check that a center default remains better than active decisions. We are given

$$\int_{-r}^r \Phi(\Delta) d\Delta \leq 2rE(c) \quad (9)$$

for $r = (\bar{s} - \underline{s})/2$, and we want to show it remains true for all lower r . The derivative of the left side with respect to r is $2\Phi(r)$, which is $\geq 2E(c)$ if $r \geq \Delta_e$. Thus, as we lower r but keep $r \geq \Delta_e$, the left side of (9) decreases faster than the right side, so (9) stays true. And if $r < \Delta_e$, then (9) holds because $\Phi(\Delta) < E(c)$ for $-r < \Delta < r$.

Next we check that a center default remains better than an offset default (if the offset default is defined): if $\bar{s} - \underline{s} \geq \Delta_e + \bar{\Delta}$ and

$$\int_{-r}^r \Phi(\Delta) d\Delta \leq \int_{-\Delta_e}^{2r-\Delta_e} \Phi(\Delta) d\Delta \quad (10)$$

for $r = (\bar{s} - \underline{s})/2$, then it also holds for all lower $r \geq (\Delta_e + \bar{\Delta})/2$. This is completely analogous to (9): for all r in the relevant range, the derivative of the left side of (10) is $2\Phi(r) \geq 2E(c)$, while the derivative of the right side is $2\Phi(2r - \Delta_e) = 2E(c)$. So as r decreases, the left side decreases faster than the right side, and (10) remains true.

This shows that if a center default is optimal, it remains optimal when the range of s is narrowed. Therefore, a center default is optimal whenever $\bar{s} - \underline{s} < w(\beta)$. By definition, a center default is not optimal for $\bar{s} - \underline{s} > w(\beta)$, so the optimal default here is either offset or active decisions.

A similar argument to the above shows that if active decisions are better than an offset default, then this remains the case when the range $\bar{s} - \underline{s}$ is increased, and similarly if offset defaults are preferred to active decisions. (As explained in the main text, the marginal employee is in the plateau of the function $\Phi(\Delta)$ in both cases.) So for fixed β , either active decision is best for all $\bar{s} - \underline{s} > w(\beta)$, or offset defaults are best for all $\bar{s} - \underline{s} > w(\beta)$. Call β an ‘‘active value’’ or an ‘‘offset value’’ accordingly. Proposition 3 implies that if some β is an active value, then all lower β are also active values.

By Proposition 5, for all sufficiently low β , active decisions are optimal for all possible distributions of s . Define β_{ac} to be the supremum of all β for which this is true when s is uniformly distributed. Part 1 of Proposition 9 immediately holds. Conversely, by

Proposition 4, when β is high enough, active decisions are never optimal. Define β_{oc} to be the supremum of all β such that active decisions are optimal for some width $\bar{s} - \underline{s}$. Thus we have $0 < \beta_{ac} \leq \beta_{oc} < 1$; moreover, all $\beta < \beta_{oc}$ are active values, and all $\beta > \beta_{oc}$ are offset values. We will check that $\beta_{ac} < \beta_{oc}$ strictly. When $\beta = \beta_{oc}$ and $\bar{s} - \underline{s} > w(\beta)$, the planner must be indifferent between offset defaults and active decisions; that is,

$$\int_{-\Delta_e}^{\bar{\Delta}} \Phi(\Delta) d\Delta = (\bar{\Delta} + \Delta_e)E(c). \quad (11)$$

Indeed, if the left side is less than the right side (so that offset defaults are preferred), then by continuity this remains true when β is lowered slightly, and active decisions are never optimal for any $\bar{s} - \underline{s}$ at this lower β . This violates the definition of β_{oc} . A similar argument applies if the left side is greater than the right side. But (11) implies that $\Phi(\Delta) < E(c)$ for some Δ , since the integral includes a hump where $\Phi(\Delta) > E(c)$. This implies that active decisions are not optimal when $\bar{s} - \underline{s}$ is sufficiently small, and this remains true when β is decreased slightly. Hence, β can be decreased slightly from β_{oc} and still be $\geq \beta_{ac}$, proving $\beta_{ac} < \beta_{oc}$.

Now, parts 2 and 3 of Proposition 9 were effectively proven two paragraphs ago. Finally, Proposition 3 implies that $\inf\{\bar{s} - \underline{s} : \text{active decisions are optimal}\}$ can never decrease when β increases. This proves part 4. ■

The last bit of our analysis is the case $\beta = 1$. This result is more general; it does not depend on the distribution of transactions costs. The point is simply that the function $\Phi(\Delta)$ consists of a valley with no humps; the optimal default always takes as much advantage as possible of this valley.

Proposition 10 *Fix κ, \underline{c} , and \bar{c} , and assume $\beta = 1$. Let $\bar{\Delta}$ be the smallest value with $\Phi(\bar{\Delta}) = E(c)$. Then*

- *if $\bar{s} - \underline{s} \leq 2\bar{\Delta}$, then a center default $d = (\bar{s} + \underline{s})/2$ is the unique optimum;*
- *otherwise, the set of optimal defaults consists of all $d \in [\underline{s} + \bar{\Delta}, \bar{s} - \bar{\Delta}]$.*

(The value of $\bar{\Delta}$ is the same as before, and it simplifies to $\sqrt{(\bar{c} - \underline{c})/2\kappa}$ when $\beta = 1$.)

Proof. First, we carry out the same analysis as in Proposition 8. When $\beta = 1$, Proposition 2 tells us that Φ is increasing in $|\Delta|$ for $|\Delta| \leq \bar{\Delta}$, and then becomes flat at $E(c)$ for $|\Delta| \geq \bar{\Delta}$.

We want to identify the d that minimizes

$$\int_{\underline{s}}^{\bar{s}} \Phi(s - d) ds.$$

The first-order condition is

$$\Phi(\underline{s} - d) = \Phi(\bar{s} - d).$$

If this common value is $< E(c)$, then by monotonicity, there are only two values $\pm\bar{\Delta}$ at which Φ takes on this value. Hence, $\underline{s} - d = -(\bar{s} - d)$, and we have a center default.

Otherwise, the common value is $E(c)$, which is $\Phi(\Delta)$ for all $\Delta \leq -\bar{\Delta}$ or $\Delta \geq \bar{\Delta}$. By symmetry we may assume $\bar{s} - d \geq \bar{\Delta}$. If $\underline{s} - d \geq \bar{\Delta}$ then we have an active-decision regime, which cannot be optimal (by Proposition 4). Hence $\underline{s} - d \leq -\bar{\Delta}$. Then the social welfare integral is equal to

$$\int_{-\bar{\Delta}}^{\bar{\Delta}} \Phi(\Delta) d\Delta + (\bar{s} - \underline{s} - 2\bar{\Delta})E(c),$$

and this value is independent of the choice of default d as long as $\underline{s} - d \leq -\bar{\Delta}$ and $\bar{s} - d \geq \bar{\Delta}$; that is, $d \in [\underline{s} + \bar{\Delta}, \bar{s} - \bar{\Delta}]$. So if one such d is optimal, all of them are.

All possible optima are of one of these two types. The first type of optimum only exists for $\bar{s} - \underline{s} < 2\bar{\Delta}$, while the second type only exists for $\bar{s} - \underline{s} \geq 2\bar{\Delta}$. The proposition follows immediately. ■

7.4 The Case of Naive Workers

In the employee's problem, analysis of naive agents is a mild extension of analysis of sophisticated agents. The naive agent expects all future selves to have short-term discount factor $\hat{\beta} > \beta$ instead of β , with all other parameters unchanged (and believes that this is common knowledge among all future selves). Therefore, she expects that all future selves will use the cutoff c^* and incur the total loss ϕ that she would have if she were a sophisticated $(\hat{\beta}, \delta)$ -discounter. We refer to these values as $\hat{c}^*, \hat{\phi}$, respectively, and continue to use c^*, ϕ to refer to the values that actually hold for the naive agent.

Proposition 11 *In the employee model, if f, β, δ , and L are held fixed, then c^* is (weakly) decreasing in $\hat{\beta}$, and ϕ is (weakly) increasing in $\hat{\beta}$.*

Proof. From Proposition 2, ϕ is decreasing in β for sophisticated agents. This implies that,

for fixed β , $\widehat{\phi}$ is decreasing in $\widehat{\beta}$. Now $\widehat{\phi}$ is the total loss that the period- t self expects to experience beginning in period $t + 1$ if she doesn't act in period t . So her total loss from not acting immediately (she believes) is $\beta\delta(L + \widehat{\phi})$. The usual indifference condition applies, and $c^* = \beta\delta(L + \widehat{\phi})$ (truncated as usual to $[\underline{c}, \bar{c}]$). Hence, c^* is decreasing in $\widehat{\beta}$.

Now, holding β fixed, let c_β^* be the cutoff used by a sophisticated (β, δ) -discounter. If the agent is naive, she will use some $c^* \leq c_\beta^*$. Her total loss ϕ is given in terms of c^* by equation (2), which does not involve β or $\widehat{\beta}$. So we just need to check that ϕ , given as a function of c^* by (2), is decreasing on $[\underline{c}, c_\beta^*]$.

Corollary 15 and Proposition 17 tell us that c^* covers $[\underline{c}, c_\beta^*]$ as β falls to 0 (for sophisticates), and Proposition 2 tells us that as β decreases, c^* decreases and ϕ increases. Since ϕ is given by (2) in the naive case too, ϕ is decreasing in c^* , as needed. ■

Proposition 13 Fix $f, \beta, \widehat{\beta}$. Assume $\delta = 1$. For all sufficiently small L , the agent will set $c^* = \underline{c}$.

Proof. Proposition 17 tells us that for a sophisticated agent, $c^* \rightarrow \underline{c}$ as $L \rightarrow 0$. So for the naive agent, $\widehat{c}^* \rightarrow \underline{c}$ as $L \rightarrow 0$. For low enough L , we have $\beta\delta(L + \widehat{\phi}) = \widehat{c}^*(\beta/\widehat{\beta}) < \underline{c}$, and the naive agent will set $c^* = \underline{c}$. ■

8 References

- Abadie, Alberto, and Sebastien Gay (2004). "The Impact of Presumed Consent Legislation on Cadaveric Organ Donation: A Cross Country Study," *Journal of Health Economics* 25(4): 599-620.
- Aguiar, Mark, and Erik Hurst (2005). "Consumption vs. Expenditure," *Journal of Political Economy* 113(5): 919-948.
- Benartzi, Shlomo, and Richard H. Thaler (2001). "Naive Diversification Strategies in Defined Contribution Saving Plans," *American Economic Review* 91(1): 79-98.
- Choi, James J., David Laibson, Brigitte C. Madrian, and Andrew Metrick (2004). "For Better or For Worse: Default Effects and 401(k) Savings Behavior," in David A. Wise, ed., *Perspectives in the Economics of Aging* (Chicago, IL: University of Chicago Press): 81-121.

- Choi, James J., David Laibson, Brigitte C. Madrian, and Andrew Metrick (2002). "Defined Contribution Pensions: Plan Rules, Participant Decisions, and the Path of Least Resistance," in James M. Poterba, ed., *Tax Policy and the Economy* 16, (Cambridge, MA: MIT Press): 67-113.
- Cronqvist, Henrik, and Richard H. Thaler (2004). "Design Choices in Privatized Social-Security Systems: Learning from the Swedish Experience," *American Economic Review* 94(2): 424-428.
- Engen, Eric M., William G. Gale, and Cori E. Uccello (1999). "The Adequacy of Household Saving," *Brookings Papers on Economic Activity* 0(2): 65-165.
- Even, William E., and David MacPherson (2005). "The Effects of Employer Matching in 401(k) Plans," *Industrial Relations* 44(3): 525-549.
- Holden, Sarah, and Jack VanDerhei (2001). "Contribution Behavior of 401(k) Plan Participants," *Investment Company Institute Perspective* 7(4).
- Hurst, Erik (2006). "Grasshoppers, Ants and Pre-Retirement Wealth: A Test of Permanent Income Consumers," University of Chicago Working Paper.
- John Hancock (2002). "Eighth Defined Contribution Plan Survey: Insight Into Participant Knowledge & Behavior." (Boston, MA: John Hancock Financial Services).
- Johnson, Eric J., and Daniel G. Goldstein (2003). "Do Defaults Save Lives?," *Science* 302(November 21): 1338-1339.
- Johnson, Eric J., John Hershey, Jacqueline Meszaros, and Howard Kunreuther (1993). "Framing, Probability Distortions, and Insurance Decisions," *Journal of Risk and Uncertainty* 7(1): 35-53.
- Johnson, Eric J., Steven Bellman, and Gerald L. Lohse (2002). "Defaults, Framing and Privacy: Why Opting In-Opting Out," *Marketing Letters* 13(1): 5-15.
- Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler (1991). "The Endowment Effect, Loss Aversion, and Status Quo Bias," *Journal of Economic Perspectives* 5(1): 193-206.

- Laibson, David (1997). “Golden Eggs and Hyperbolic Discounting,” *Quarterly Journal of Economics* 112(2): 443-477.
- Laibson, David, Andrea Repetto, and Jeremy Tobacman (2007) “Estimating Discount Functions with Consumption Choices Over the Lifecycle.” NBER working paper 13314.
- Madrian, Brigitte C., and Dennis Shea (2001). “The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior,” *Quarterly Journal of Economics* 116(4): 1149-1187.
- O’Donoghue, Ted, and Matthew Rabin (1999a). “Procrastination in Preparing for Retirement,” in Henry J. Aaron, ed., *Behavioral Dimensions of Retirement Economics* (Washington, D.C.: Brookings Institution Press; New York, NY: Russell Sage Foundation): 125-156.
- O’Donoghue, Ted, and Matthew Rabin (1999b). “Doing It Now or Later,” *American Economic Review* 89(1): 103-124.
- Park, C. Whan, Sung Y. Jun, and Deborah J. MacInnis (2000). “Choosing What I Want Versus Rejecting What I Do Not Want: An Application of Decision Framing to Product Option Choice Decisions,” *Journal of Marketing Research* 37(2): 187-202.
- Phelps, Edmund S., and Robert A. Pollak (1968). “On Second-Best National Savings and Game-Equilibrium Growth,” *Review of Economic Studies* 35(2): 185-199.
- Profit Sharing/401(k) Council of America (2001). “Automatic Enrollment 2001: A Study of Automatic Enrollment Practices in 401(k) Plans.” Downloaded from <http://www.pasca.org/data/autoenroll2001.asp> on April 19, 2001.
- Samuelson, W., and R. J. Zeckhauser (1988). “Status Quo Bias in Decision Making,” *Journal of Risk and Uncertainty* 1: 7-59.
- Sunstein, Cass R., and Richard H. Thaler (2003). “Libertarian Paternalism,” *American Economic Review* 93(2): 175-179.
- Thaler, Richard H., and Shlomo Benartzi (2004). “Save More TomorrowTM: Using Behavioral Economics to Increase Employee Saving,” *Journal of Political Economy* 112(S1): S164-S187.

Warshawsky, Mark J., and John Ameriks (2000). "How Prepared Are Americans for Retirement?" in Olivia S. Mitchell, P. Brett Hammond, and Anna M. Rappaport, eds., *Forecasting Retirement Needs and Retirement Wealth* (Philadelphia, PA: University of Pennsylvania Press): 33-67.

Table 1. 401(k) plan features by effective date

	Effective January 1, 1997	Effective November 23, 1997
<i>Eligibility</i>		
Eligible employees	U.S. employees, age 18+	U.S. employees, age 18+
First eligible	Full-time employees eligible upon hire; part-time employees must accrue 1,000 hours in 1 year	Full-time employees eligible upon hire; part-time employees must accrue 1,000 hours in 1 year
Employer match eligible	Immediately upon plan eligibility	Immediately upon plan eligibility
<i>Enrollment</i>		
	First 30 days of employment or January 1 of succeeding calendar years	Daily
<i>Contributions</i>		
Employee contributions ^a	Up to 17% of compensation	Up to 17% of compensation
Non-discretionary employer match rate	50% of employee contribution	50% of employee contribution
Discretionary employer match rate	Up to 100% (50% for bonus-eligible employees); rate depended on company profitability	Up to 100% (50% for bonus-eligible employees); rate depended on company profitability ^b
Employer match restrictions	Match on the lesser of before-tax employee contribution or 5% of compensation; match invested in employer stock	Match on the lesser of before-tax employee contribution or 5% of compensation; match invested in employer stock
<i>Employer match vesting</i>		
	Immediate	Immediate
<i>Other</i>		
Loans	Not available	Available; 2 maximum
Hardship withdrawals	Available	Available
Investment choices	4 options; employer stock also available, but only for after-tax contributions and employer match	6 options + employer stock (available for before- and after-tax contributions)

^aTotal employee contributions within each year were capped by federal law at \$9,500 (1997), \$10,000 (1998-99), and \$10,500 (2000-01).

^bActual discretionary match rates were 20% (1995), 20% (1996), 100% (1997), 100% (1998), 27% (1999), 33% (2000), 0% (2001).

Table 2. Comparison of worker characteristics

	Study company			U.S. workforce (3/98 CPS)
	Active-decision cohort on 12/31/98	Standard- enroll. cohort on 12/31/99	All workers on 12/31/99	
<i>Average age (years)</i>	34.7	34.1	41.0	38.8
<i>Gender</i>				
Male	47.6%	42.0%	45.0%	53.1%
Female	52.4%	58.0%	55.0%	46.9%
<i>Marital Status</i>				
Single	41.4%	49.3%	32.4%	39.0%
Married	56.0%	49.7%	67.6%	61.0%
<i>Compensation</i>				
Avg. monthly base pay	\$3,043	\$2,869	\$4,451	--
Median monthly base pay	\$2,666	\$2,513	\$3,669	--
Avg. annual income ^a	\$35,381	\$33,197	\$51,792	\$32,414
Median annual income ^a	\$31,013	\$29,239	\$41,190	\$24,108
<i>Geography</i>				
East	13.1%	11.1%	15.3%	18.9%
Midwest	34.3%	37.6%	32.2%	24.1%
South	37.7%	38.9%	37.8%	34.7%
West	14.7%	12.3%	14.7%	22.4%
<i>Number of Employees</i>	2,231	2,349	46,822	--

The samples in the first three columns are taken from individuals employed at the study company as of the dates indicated in the column title. The sample in the last column is all individuals in the March 1998 Current Population Survey who worked in the previous year (weighted). Compensation is in 1998 dollars. Figures may not add up to 100% because of missing data and employees located in Puerto Rico.

^aThe annual-income measure that is reported to us for the study company is the employee's annual taxable (W2) income. Annual income for the U.S. workforce calculated from the CPS is total annual labor earnings in the previous calendar year, some of which may be non-taxable.

Table 3. Tobit regression of contribution rates under two 401(k) enrollment regimes

Intercept	2.915**
	(0.191)
<i>Female</i>	0.545
	(0.403)
<i>Married</i>	1.335**
	(0.383)
$\text{Log}(\text{Base pay})$	4.898**
	(0.666)
$0 \leq \text{Age} < 30$	-6.829*
	(2.681)
$30 \leq \text{Age} < 40$	-5.144
	(2.683)
$40 \leq \text{Age} < 50$	-5.133
	(2.703)
$50 \leq \text{Age} < 60$	-3.182
	(2.753)
<i>Active-decision cohort</i>	0.086
	(0.247)
<i>Active decision cohort</i> \times <i>Female</i>	-1.989**
	(0.547)
<i>Active-decision cohort</i> \times <i>Married</i>	-0.528
	(0.503)
<i>Active-decision cohort</i> \times $\text{Log}(\text{Base pay})$	-1.930
	(1.053)
<i>Active-decision cohort</i> \times ($0 \leq \text{Age} < 30$)	-0.584
	(3.553)
<i>Active-decision cohort</i> \times ($30 \leq \text{Age} < 40$)	-0.577
	(3.552)
<i>Active-decision cohort</i> \times ($40 \leq \text{Age} < 50$)	0.004
	(3.578)
<i>Active-decision cohort</i> \times ($50 \leq \text{Age} < 60$)	-0.103
	(3.659)
$\ln(\sigma_{AD}/\sigma_{SE})$	-0.137**
	(0.042)
N	3,488

If the employee is in the active-decision cohort, the dependent variable is the 401(k) contribution rate (in percentage points) 3 months after hire; if the employee is in the standard-enrollment cohort, the dependent variable is the contribution rate 30 months after hire. Independent variables are log of base pay, a dummy for being in the active-decision cohort, and gender, marital status, and age range dummies, calculated as of the contribution-rate date. Demographic variables are de-measured. Both cohorts are restricted to employees who remain in the data for at least 30 months. The tobit regression assumes that errors are normal and homoskedastic within each cohort but possibly heteroskedastic across cohorts. Robust standard errors are reported in parentheses under the point estimates.

*Significant the 5% level **Significant at the 1% level

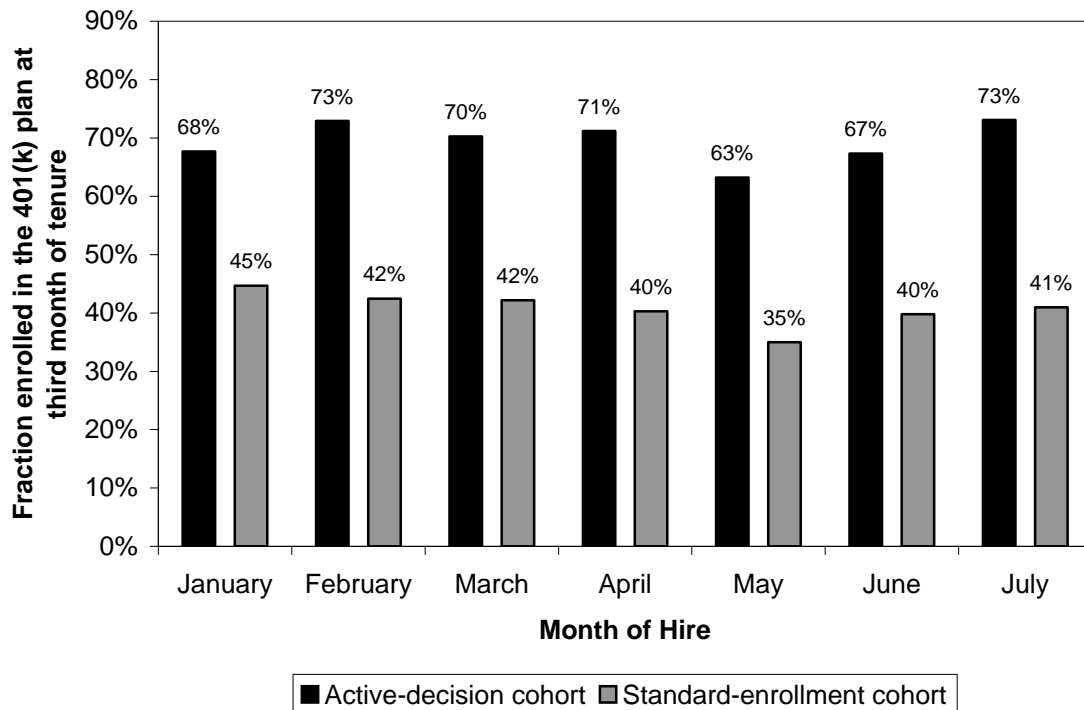


Figure 1. Fraction of employees enrolled in the 401(k), by hire month. The fraction displayed is as of the third month of tenure at the company. The active-decision cohort was hired between January and July 1997. The standard-enrollment cohort was hired between January and July 1998.

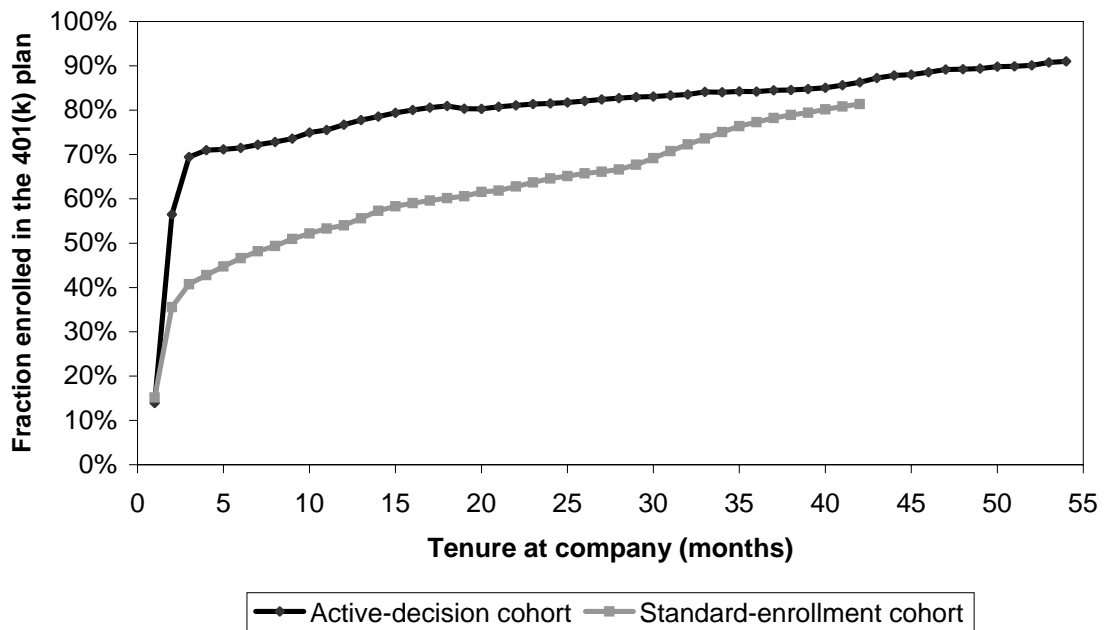


Figure 2. Fraction of employees enrolled in the 401(k) plan, by tenure at company. An employee is counted enrolled in the 401(k) even if he or she has stopped contributing to the plan. The series are not monotonically rising because they are constructed from multiple cross-sections, so the samples are not fixed over time.

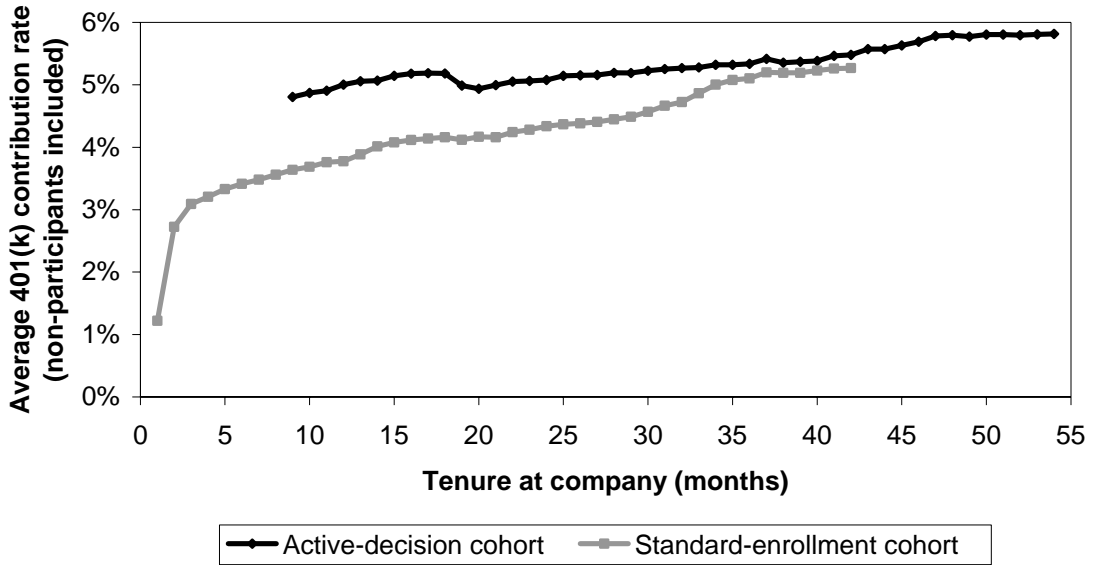


Figure 3. Average 401(k) contribution rate by tenure at company. At each point, the averages include employees not currently contributing to the 401(k) plan; their contribution rate is zero.

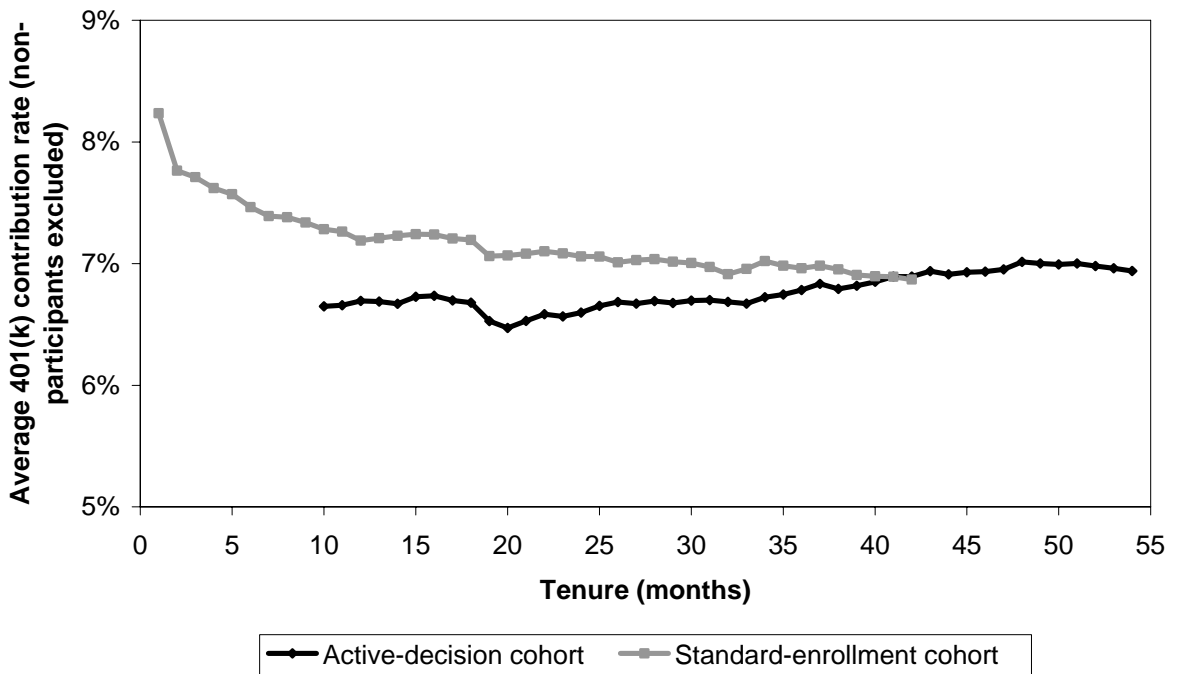
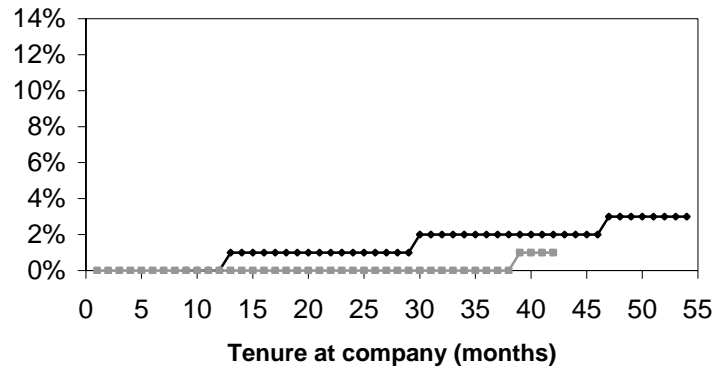
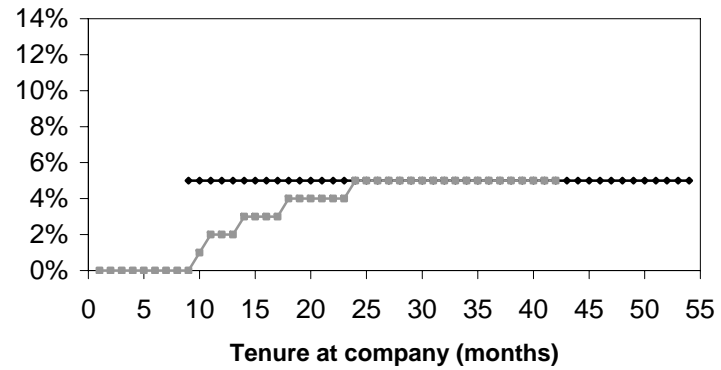


Figure 4. Average 401(k) contribution rate among 401(k) participants by tenure at company. At each point, the averages exclude employees not currently contributing a positive amount to the 401(k) plan.

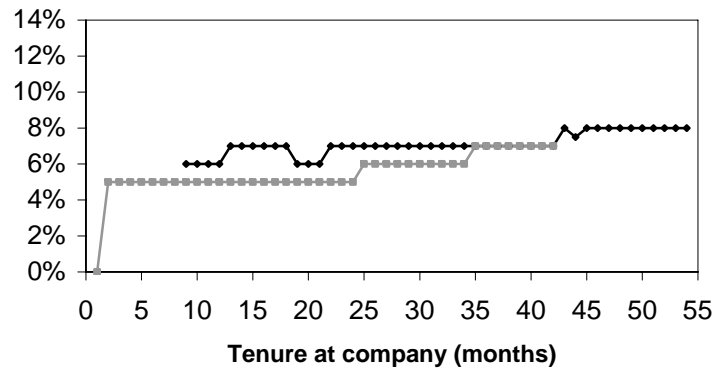
25th percentile contribution rate



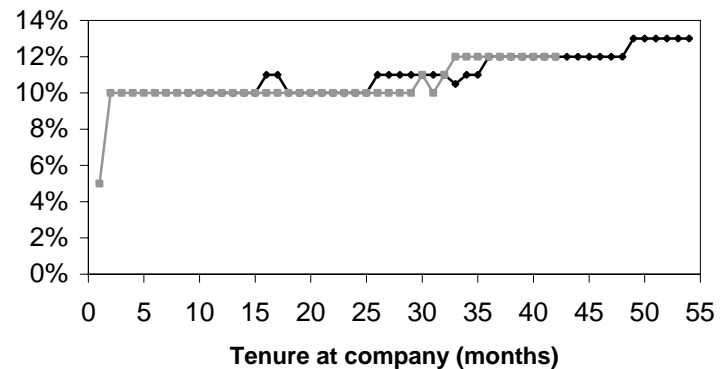
50th percentile contribution rate



75th percentile contribution rate



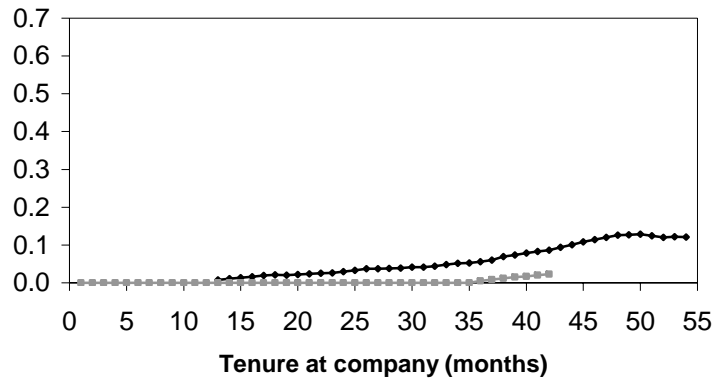
90th percentile contribution rate



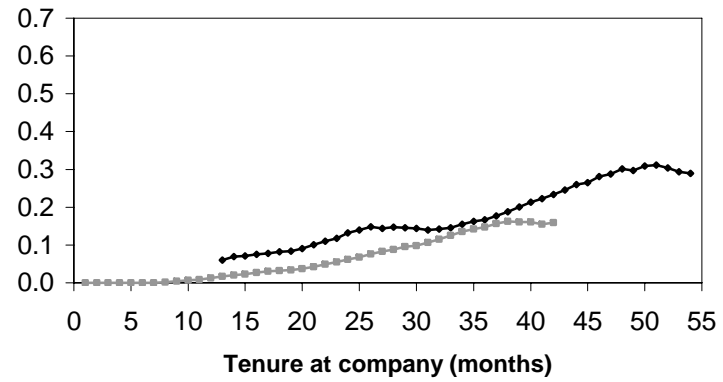
—◆— Active-decision cohort - - - ■ - - - Standard-enrollment cohort

Figure 5. 401(k) contribution rates at different contribution-rate percentiles. The percentile breakpoints are calculated separately for each cohort at each point in time.

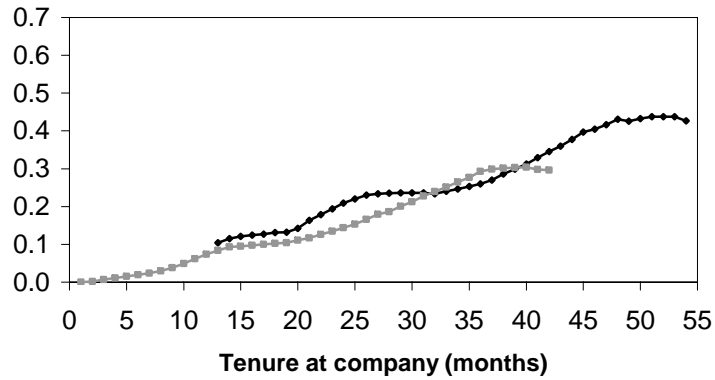
25th percentile balance-to-base pay ratio



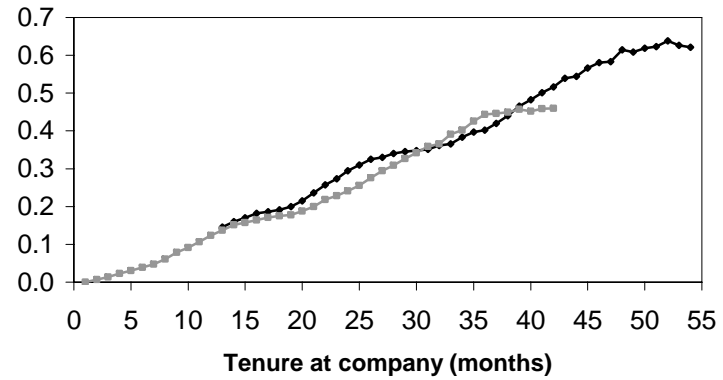
50th percentile balance-to-base pay ratio



75th percentile balance-to-base pay ratio



90th percentile balance-to-base pay ratio



—◆— Active-decision cohort —■— Standard-enrollment cohort

Figure 6. 401(k) balance-to-base pay ratios at different balance-to-base pay percentiles. The balances exclude outstanding loan principal and any money rolled into the account from a former employer. The percentile breakpoints are calculated separately for each cohort at each point in time.

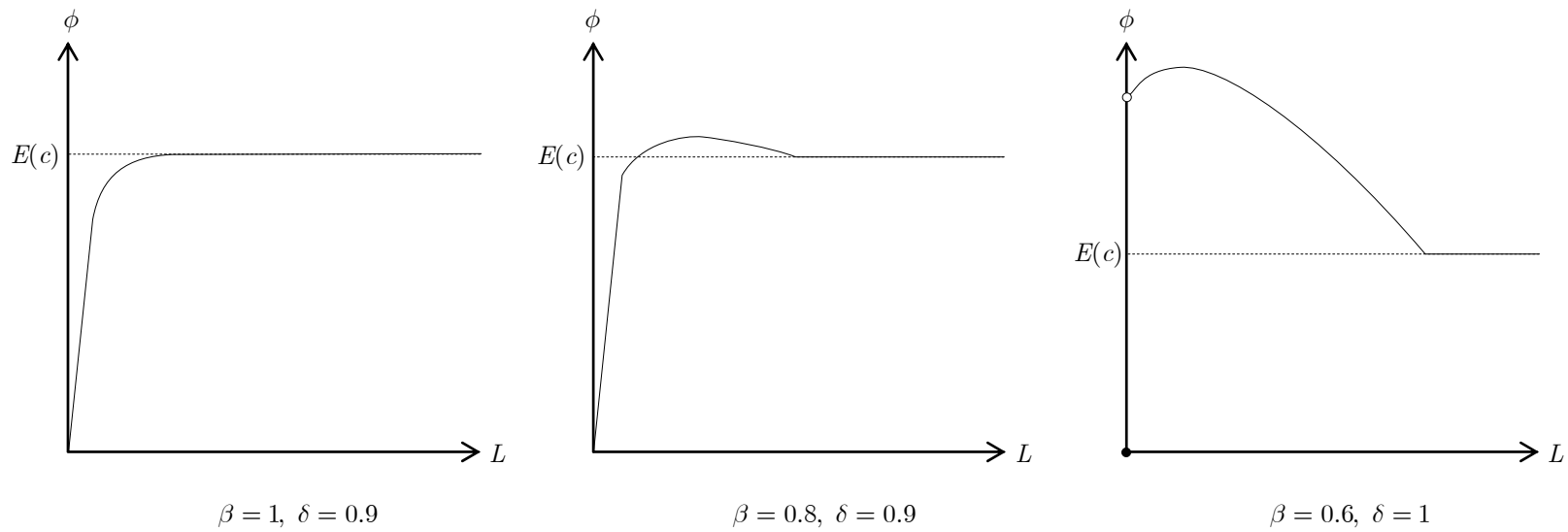


Figure 7. Employee's total expected loss ϕ as a function of per-period flow loss L from not opting out. The parameters specific to each panel, the quasi-hyperbolic discount factor β and the exponential discount factor δ , are shown beneath each graph. In all panels, opt-out costs are assumed to be uniformly distributed between 0.75 and 1.25. The left and center panels have the same y -axis scale, but the right panel has a different y -axis scale.

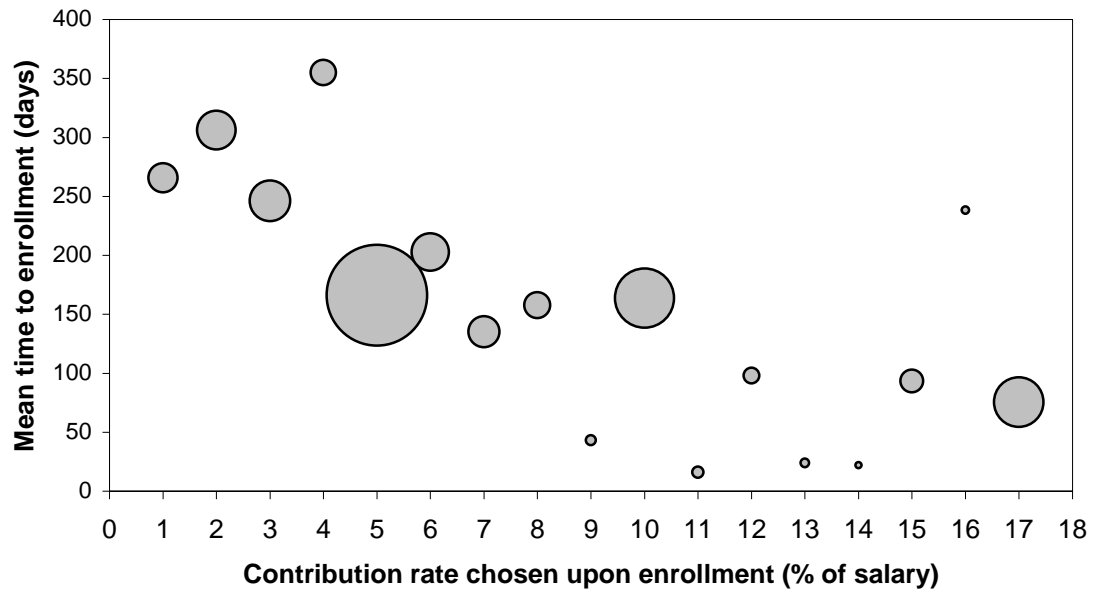


Figure 8. Mean time between hire date and enrollment by contribution rate in the standard-enrollment cohort. The area of each bubble is proportional to the number of employees it represents. The sample consists of all employees in the standard-enrollment cohort who worked at the company for at least 30 months and enrolled within 30 months of hire.

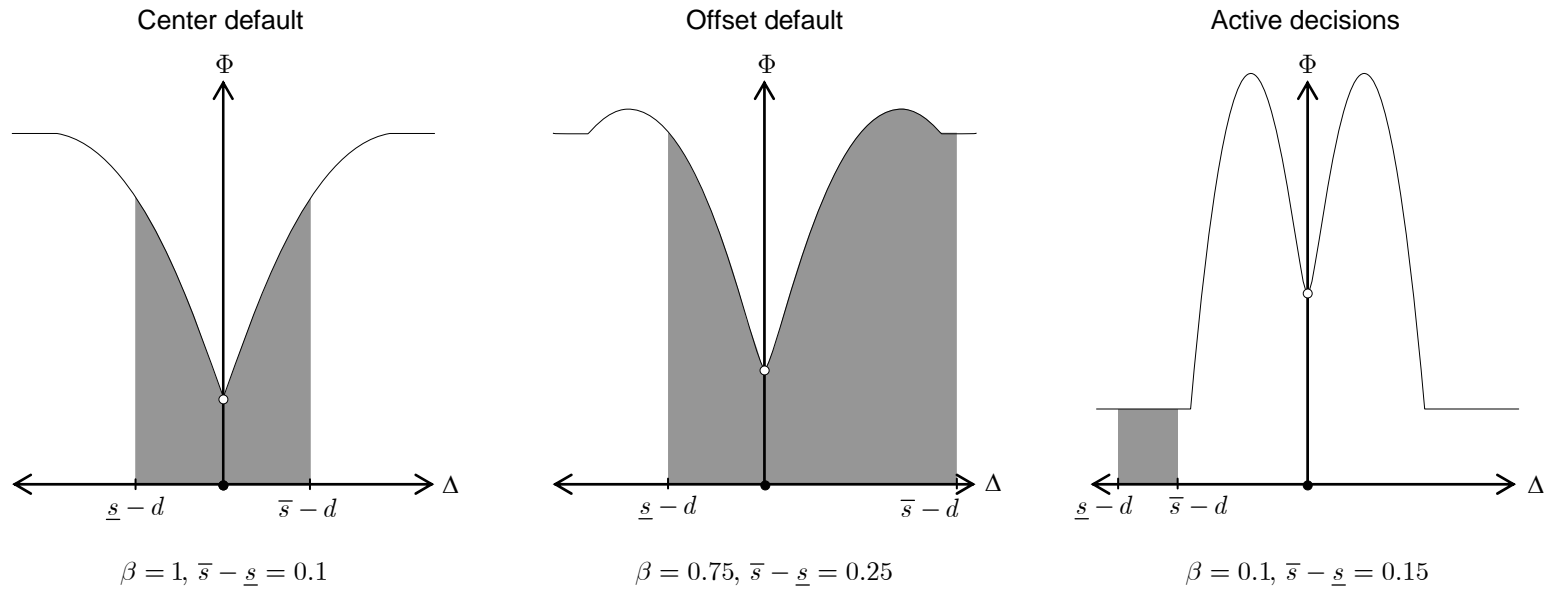


Figure 9. Possible optimal default regimes when opt-out costs and optimal savings rates are uniformly distributed and $\delta = 1$. The panels illustrate parameter values that support the three classes of optimal defaults: the center default, the offset default, and active decisions. The shaded area in each panel represents the social welfare losses generated by the corresponding default regime. The parameters specific to each panel, the quasi-hyperbolic discount factor β and the range of optimal savings rates $\bar{s} - \underline{s}$, appear below each figure. In all three panels, the opt-out cost is uniformly distributed between 0.25 and 1.75, and the loss function scaling factor $\kappa = 100$. The left and center panels have the same y -axis scale, but the right panel has a different y -axis scale.

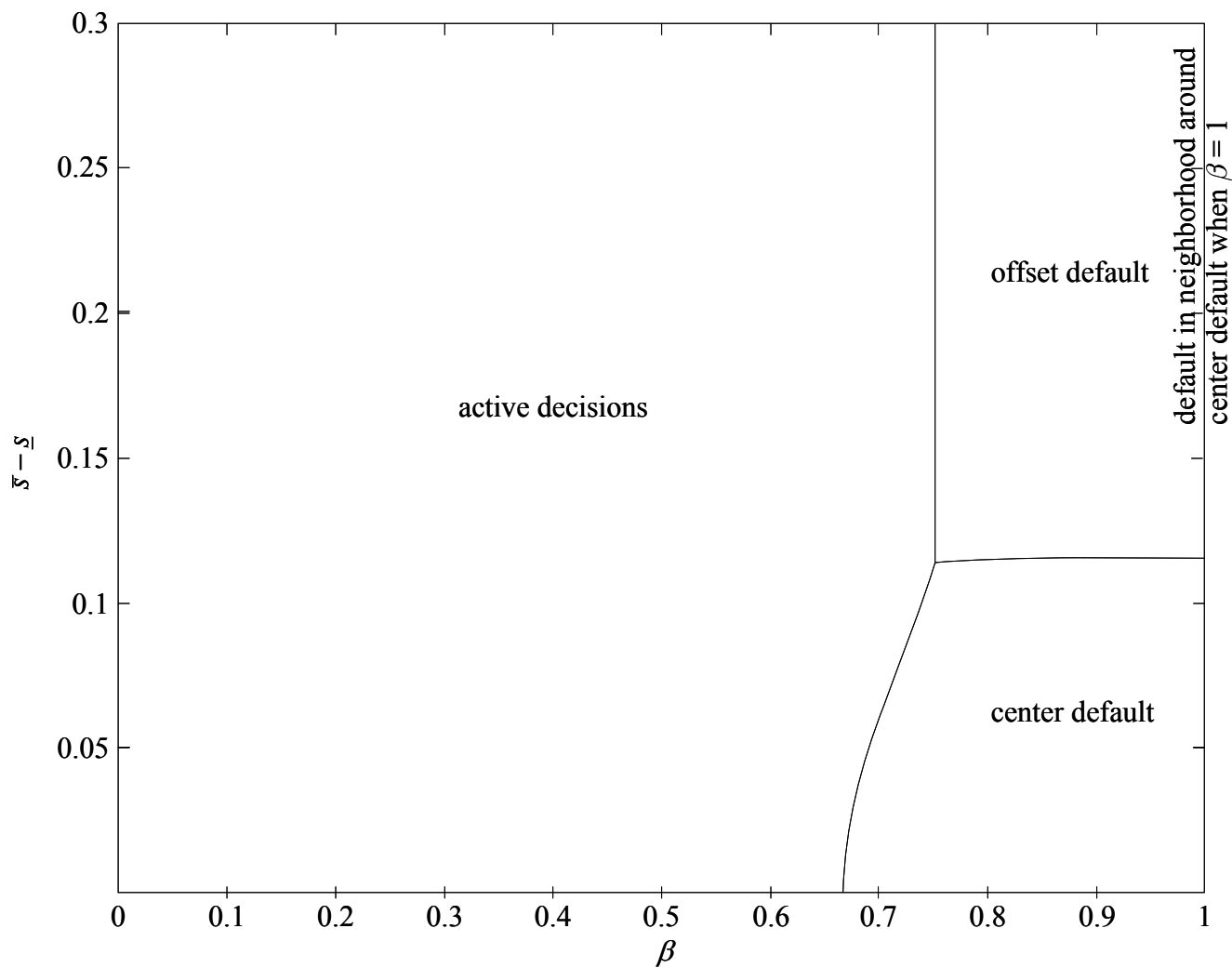


Figure 10. Characterization of optimal default regimes when opt-out costs and optimal savings rates are uniformly distributed and $\delta = 1$. This figure shows the boundaries of the optimal default regimes as a function of the quasi-hyperbolic discount factor β and the range of optimal savings rates $\bar{s} - \underline{s}$ when the opt-out cost is uniformly distributed between $2/3$ and $4/3$, and the loss function scaling factor $\kappa = 100$.