A POSITIVE THEORY OF RETIREMENT PLAN DESIGN

RYAN BUBB* AND PATRICK L. WARREN**

ABSTRACT. We develop a positive theory of employer-sponsored retirement plan design using a behavioral contract theory approach. Equilibrium in the labor market results in retirement plans that generally cater to, rather than correct, workers' mistakes. Naive myopic workers, who overestimate their future savings, are offered matching contributions, which can help offset their present bias but result in harmful cross-subsidization of rational workers. Workers who are sensitive to defaults are offered automatic enrollment (opt-out) plans, but only when the savings-increasing effect on procrastinators is outweighed by the savings-decreasing effect on those who interpret defaults as implicit advice. The investment options in employer plans include low-cost funds, for rational workers, but also include high-cost alternatives when some workers diversify naively. Our theory provides novel explanations for a range of facts about retirement plan design and calls into question the practice of depending on employers to design plans to counteract the mistakes of workers.

1. INTRODUCTION

Employer-sponsored retirement savings plans are the predominant vehicle for private retirement savings in the United States, in large part due to their preferential tax treatment. A growing empirical literature shows that the design of these plans affects savings behavior in ways inconsistent with rational optimization. For example, the default rule that governs plan participation—an inconsequential detail of plan design from the perspective of neoclassical economic theory—has substantial effects on worker savings behavior (Madrian and Shea, 2001).

These empirical findings have also informed normative claims by behavioral economists about how employers *should* design their retirement savings plans. In a survey article, Benartzi and Thaler (2007) ask, "What can employers do so that more plan participants enroll in retirement plans, contribute an amount that will build a reasonable retirement nest-egg, and allocate the funds among assets in an appropriately diversified way?" They proceed to suggest to employers a range of plan design options to improve their workers' retirement savings outcomes. Employers should paternalistically harness the stickiness of default rules, for example, to counteract myopic workers' temptation to save too little (Thaler and Benartzi, 2004; Carroll, Choi, Laibson, Madrian, and Metrick, 2009). These papers take a "public finance" approach to retirement plan design, casting the employer in the role of a social planner, designing its retirement plan to maximize social welfare.

But despite the considerable behavioral literature on retirement savings plans, there has been virtually no work on a basic positive question: what *incentives* do employers have in designing their retirement plans for biased workers? In this paper we take a behavioral contract theory approach to studying this question. Retirement plans are an important feature of compensation contracts in the labor market, designed by employers to attract workers. We analyze the equilibrium plan designs that emerge when employers contract with biased workers. We focus on two key dimensions of retirement plan design: (1) the rules that structure and incentivize contributions to the plan; and (2) the set of investment options available within the plan.

We begin in Section 2 with the rules governing plan contributions. The majority of employer-sponsored retirement plans today are defined contribution plans, which are funded by a mix of

employer contributions and employee contributions. The employer contributions commonly come in two forms: non-elective contributions and matching contributions. Another important plan design feature is the default rule governing employee contributions. In recent years many employers have adopted plans in which new hires are automatically enrolled at some positive contribution rate unless they affirmatively opt out. This shift has been a response to legislative changes—enacted by Congress at the urging of behavioral economists—that eliminated regulatory barriers to using automatic enrollment. The hope behind this approach was that employers would then adopt defaults that would increase retirement savings by pointing workers in a pro-saving direction when they fail to make active decisions on their own (Orszag, Iwry, and Gale, 2006).

To analyze employer incentives for choosing these rules that structure contributions to the plan, we assume that a set of identical firms compete over workers by offering compensation packages consisting of a wage and a retirement plan. In our baseline model, retirement plans are composed of an employer non-elective contribution and a matching contribution of m dollars for every dollar the worker saves. Workers evaluate firms' contract offers with an eye toward securing consumption both during their working life and during retirement. We start by considering homogenous workers with given intertemporal preferences. The neoclassical benchmark entails a simple wage contract. Retirement plans serve no useful purpose for rational, time-consistent exponential discounters (tax benefits aside), and matching contributions are actively harmful by distorting rational workers' intertemporal consumption choices.

Retirement plans emerge in equilibrium, however, if instead workers suffer from some degree of present bias or myopia, which would generally lead them to save less for retirement than they would prefer ex ante. Sophisticated myopic workers understand their self-control problem. The problem for a sophisticated worker is to choose a contract with a retirement plan that induces her future self to save optimally. This can be achieved through non-elective employer contributions, locking in some level of retirement savings, or matching contributions, counterbalancing her present bias. Sophisticated myopic workers in equilibrium receive a retirement plan that provides a first-best commitment device.

Imperfectly sophisticated myopic workers, in contrast, underestimate their degree of present bias and hence their need for commitment devices. But we show that such naive workers have an additional motivation for choosing a contract that offers matching contributions: they overestimate how much they will save and therefore the amount of matching contributions they will receive. This overestimation motivation does not, however, result in a level of matching that is finely calibrated to naives' need for commitment. We show that the equilibrium level of matching for naives is a function of their elasticity of intertemporal substitution and can either overshoot or undershoot the first-best.

We also analyze the case with heterogeneous types. Naive myopic workers pool with rational workers in matching contracts, which results in cross-subsidization of rationals by naives since rationals receive more matching contributions. This cross-subsidization harms naives both directly by lowering their total compensation and indirectly by distorting the equilibrium matching rate. We show that the use of caps on the employer match facilitates the pooling of naive and rational workers in matching contracts.

We extend the model in Section 2.5 to analyze employers' choice of the default employee contribution rate. When some workers can be influenced by the default, employers have incentives to choose the default that *minimizes*, rather than raises, worker savings given the other terms of contract. The reason is that the default is not salient to workers at the time of contracting, and in equilibrium firms choose savings-minimizing defaults in order to lower the level of matching contributions they must make. Doing so relaxes their zero-profit constraint and thus allows them to offer better terms on the salient dimensions of compensation.

Whether a positive default contribution rate (i.e., automatic enrollment) results in lower savings in the equilibrium contract than a zero default (i.e., opt-in) depends on the parameters. The model predicts that employers will adopt automatic enrollment if and only if the resulting reduction in savings by those who follow defaults as implicit advice is larger than the increase in savings by procrastinators who would not have opted in. If the employer does automatically enroll employees, it will set the default contribution amount below the contract's cap on employee contributions that the employer matches.

Our theory's predictions line up well with many key facts about employer retirement plan design. The vast majority of defined contribution plans offer matching contributions, and a substantial fraction of workers in such plans fail to contribute enough to receive the full match. Moreover, most employers that have adopted automatic enrollment have chosen the minimum default initial contribution rate allowed under the regulatory safe-harbor Congress created for such plans. In the vast majority of automatic enrollment plans, the default is set below the amount needed to receive the full employer match. Existing evidence suggests that employer adoption of automatic enrollment has failed to raise, and may even have lowered, overall retirement savings (Bubb and Pildes, 2014).

We turn in Section 3 to the second main aspect of retirement plan design: the set of investment options available in the plan. Most defined contribution plans provide a menu of mutual funds and other investment options in which employees can invest their plan funds. The contractual arrangements in the industry are such that employers receive discounts on the costs of administering their retirement plan that are financed by the fees charged by the investment managers of the plan's fund options. Existing empirical evidence shows that workers make various mistakes in constructing their retirement portfolio. For example, many engage in "naive diversification" by allocating their savings equally among all of the investment options in their retirement plan (Benartzi and Thaler, 2001).

We show that when employers contract with a mix of rational workers and naive diversifiers, in equilibrium each employer's retirement plan offers a set of investment options that includes a low-fee option (for rationals) and higher-fee options (for naives) that are otherwise equivalent. Because high investment fees are not salient to naive diversifiers, and are avoidable by workers for whom they are salient, including high-cost investment options lowers the cost to the employer of offering the retirement plan and thereby increases the salient dimensions of workers' compensation. Our predictions match recent evidence showing that the set of investment options in most employer plans includes so-called "dominated funds": high-cost options that have an optimal portfolio weight of less than 1% (Ayres and Curtis, 2015).

The approach we take to analyzing employer-sponsored retirement plans builds on an existing literature in behavioral contract theory that so far has focused on firms' product markets, such as consumer credit (DellaVigna and Malmendier, 2004; Heidhues and Kőszegi, 2010; Bar-Gill, 2012), cell-phone service (Grubb, 2009), add-on goods (Gabaix and Laibson, 2006), and gym memberships (DellaVigna and Malmendier, 2006). An important theme in this literature is that the equilibrium contracts can be "exploitative" in the sense that a central consideration driving their design is an attempt by firms to profit from consumers' mistakes (Kőszegi, 2014).

One potential justification for not taking a similar approach to understanding employer-sponsored retirement plans is the view that markets do not provide important incentives for employers with respect to retirement plan design. For example, Barr, Mullainathan, and Shafir (2013) argue that attempts to boost participation in retirement plans face "at worst indifferent and at best positively inclined employers and financial firms." They contrast this with other markets, like consumer credit, in which firms have strong incentives to exploit consumer mistakes. But as we show in this paper, the behavioral contract theory approach produces a rich positive theory of employer-sponsored retirement plan design.

An important implication of our analysis is that recent attempts by behavioral economists to reform employer-sponsored retirement plans by simply showing employers what plan designs would improve worker savings outcomes and removing regulatory barriers to offering them (see, e.g., Thaler and Benartzi, 2004; Orszag, Iwry, and Gale, 2006) are unlikely to be effective. In our model, even if employers were paternalistically motivated, competition in the labor market leaves no room for paternalistic employer preferences to be expressed. Any employer that offers a zero-profit labor contract that corrects rather than caters to worker mistakes would not attract any workers. And any contract that both attracts workers and corrects their mistakes would earn negative profits and therefore not be economically viable. Conversely, if a firm has market power, then so long as it is a profit maximizer, the basic results from our model with perfect competition would hold. A profit-maximizing monopolist would have incentives to cater to worker biases in order to reduce the cost of labor. Meaningful employer paternalism in retirement plan design thus requires

a concurrence of two factors that is unlikely to be widespread: employers must have *both* paternalistic preferences and significant market power. If the motivation for retirement savings policy generally, and the preferential tax treatment of employer plans specifically, is to correct mistakes workers make in planning and saving for retirement (Kotlikoff, 1987), then our analysis shows that the delegation of plan design to employers will result in perverse outcomes for the myopic and inertial workers that retirement savings policy aims to help.

2. Plan Contributions

The way in which defined contribution retirement plans are funded is a central aspect of plan design. We begin our analysis by focusing on employer contributions. The baseline model in Section 2.2 considers homogenous worker types with varying degrees of myopia and a contract space composed of a wage and a retirement plan funded by a non-elective employer contribution and a simple linear matching contribution. In Section 2.3 we allow for heterogenous worker types. We introduce caps on employers' matching contributions in Section 2.4. In Section 2.5 we enrich the behavioral type space to include passive savings behavior and enlarge the contract space to include a default rule governing employee contributions. We conclude our analysis of plan contributions in Section 2.6 by considering how well the implications of the model fit key facts about employer-sponsored retirement plan design.

2.1. **Setup.** Consider a perfectly competitive labor market.[1] Labor contracts specify a wage $w \geq 0$ and a retirement plan that is composed of a non-elective employer contribution to the plan, $r \geq 0$, plus employer matching contributions of $m \geq 0$ dollars for every dollar the worker saves for retirement. Total employer retirement plan contributions are thus $r + sm$, where $s \geq 0$ is the amount the worker voluntarily chooses to save for retirement. We model retirement plans in this way to match the basic structure of employer-sponsored retirement plans we observe in the real

---

[1]The basic pressures we identify here would also obtain in other more monopsonisitic markets, as monopsonistic firms would still like to find the minimum-cost manner of providing a given level of perceived compensation.

world.[2] Profits from an employed worker are given by $\pi = \gamma - w - sm - r$, where γ is the value the worker produces.

Workers have access to a savings technology through their employer's retirement savings plan with a rate of return normalized to zero, but they cannot borrow. For simplicity we have assumed away any motivation to save outside of the employer's retirement plan. There is thus no need to incorporate taxes to reflect the favorable tax treatment of employer-sponsored retirement plans relative to other forms of savings.

There are three periods in which the sequence of decisions is as follows.

- Period 0: Firms make contract offers (w, r, m) and workers choose among offers.
- Period 1: Workers receive wage w and decide how much of the wage to save, s, consuming the remainder, $w - s$.
- Period 2: Retired workers consume their savings and retirement plan benefits, $(1+m)s+r$.

A worker's period-0 self ("self 0") has utility $u(c_1) + u(c_2)$, where c_i is anticipated consumption in period i, $u(\cdot)$ is increasing and concave, and the discount factor is normalized to one.

Self 1, by contrast, chooses savings to maximize the utility function $u(c_1) + \beta u(c_2)$, where $\beta \in (0, 1]$ is the worker's time-inconsistent present-bias factor. Thus, facing a contract (w, r, m), self 1 solves,

$$(1) \qquad \max_{s \geq 0} u(w - s) + \beta u(r + (1 + m)s).$$

If $\beta < 1$, we refer to the worker as myopic. If $\beta = 1$, we refer to the worker as rational.

Self 0 chooses a contract to maximize her utility taking into account her anticipated future savings behavior. Importantly, however, we assume that self 0 *believes* that self 1 will choose savings by applying a present bias factor $\hat{\beta} \in [\beta, 1]$, following O'Donoghue and Rabin (2001)'s

---

[2]We could model retirement plan contracts in a more general way by allowing firms to offer choice sets over consumption paths. But that approach produces equilibrium contracts in which workers' compensation is discontinuous in their savings behavior in a way that does not match real-world compensation contracts (see Heidhues and Kőszegi (2010) for a version of this result in the credit market context). One potential reason for the failure of that modeling approach to predict the retirement plan structures we actually observe is that, under that approach, even small mistakes by firms in assessing the savings preferences of workers (or other factors that affect savings behavior) would lead them to suffer large losses through excessive compensation.

approach to modeling partial naivete. We refer to myopic workers with $\hat{\beta} = \beta$ as sophisticated, and to those with $\hat{\beta} > \beta$ as naive.

2.2. Equilibrium Contracts with Homogenous Types.

We begin by assuming homogenous workers of a single type $(\beta, \hat{\beta})$. This can also be thought of as the case in which firms observe workers' types so that each type gets its own contract. We consider heterogenous types in Section 2.3 below.

2.2.1. *The constrained optimization problem.*

Firms are willing to offer any contract that would result in nonnegative profits, given workers' actual savings behavior, but perfect competition implies that firms must break even in equilibrium. Equilibrium labor contracts are the zero-profit contracts that maximize self 0's utility, given her beliefs about self 1's savings behavior. They are thus the solution to,

$$(2) \qquad \max_{w,r,m} u(w - s(w, r, m|\hat{\beta})) + u(r + (1 + m)s(w, r, m|\hat{\beta})),$$

subject to,

$$(3) \qquad w + r + ms(w, r, m|\beta) = \gamma,$$

$$(4) \qquad s(w, r, m|\hat{\beta}) = \arg\max_{s \geq 0} u(w - s) + \hat{\beta}u(r + (1 + m)s),$$

$$(5) \qquad s(w, r, m|\beta) = \arg\max_{s \geq 0} u(w - s) + \beta u(r + (1 + m)s).$$

Self 0 wants to maximize the sum of her utility from consumption in the two periods, as reflected in the objective function in (2). The zero-profit constraint (3) provides that total compensation paid across the two periods must equal the worker's product γ. By concavity of the utility function, the first-best outcome equates consumption in each of the two periods at $\gamma/2$.

Self 0 chooses a contract based on her belief that self 1 will put a present-bias factor of $\hat{\beta}$ on second-period utility when choosing how much to save under the contract; her anticipated savings

8

level is determined by (4). Her self 1 will actually make savings decisions according to (5), using a present-bias factor of $\beta \leq \hat{\beta}$.

2.2.2. *Equilibrium contracts.* We begin our analysis by building some basic economic intuitions for the problem. Consider first a sophisticated myopic worker. A sophisticated worker's self 0's beliefs about her self 1's savings are correct, since $\hat{\beta} = \beta$. The problem for a sophisticated worker's self 0 is to choose a contract that induces her present-biased self 1 to save optimally. It is easy to see that a sophisticated worker will be willing to choose $r = w = \gamma/2$ to solve her time-inconsistency problem through r and achieve the first best. This contract will give self 1 exactly what self 0 wants her to consume. Self 1 will want to consume even more than $\gamma/2$ in the first period, but the remaining $\gamma/2$ of her compensation is only paid in the second period through r.

A sophisticated worker can also achieve the first-best through m. The first-order condition for self 1's choice of savings in (5) is:

$$(6) \qquad -u'\big(w - s(w, r, m|\beta)\big) + \beta(1 + m)u'\big(r + (1 + m)s(w, r, m|\beta)\big) = 0.$$

Inspection of this first-order condition shows that choosing m such that $1 + m = 1/\beta$ will perfectly counterbalance self 1's present-bias, inducing self 1 to make savings decisions according to self 0's preferences, i.e., to equate her consumption in the two periods. Denote this m as $m^{FB} \equiv \frac{1 - \beta}{\beta}$.

A naive worker's self 0, in contrast, does not fully anticipate the degree of her myopia. She believes that her self 1 will save more than she actually will, since $\hat{\beta} > \beta$. So long as $\hat{\beta} < 1$, she will understand that she suffers from some degree of time-inconsistency and hence will have some degree of commitment motivation for using r and m. But a naive worker has an additional motivation for using m: she overestimates the value of m due to mistakes in projecting her future savings. A zero-profit contract that offers some $m > 0$ and induces some anticipated savings $s > 0$ will appear to the naive worker to offer total compensation greater than γ since she overestimates how much she will save under the match. It will in fact induce savings that result in total compensation of only γ. So even a completely naive worker, with $\hat{\beta} = 1$, who has no awareness of

9

her time-inconsistency and therefore no demand for commitment devices per se, will nonetheless demand some amount of matching contributions due to this overestimation mistake.

With intuitions from our preliminary analysis in place, we now proceed to characterize the equilibrium. The contract space for the problem in (2) - (5) contains two commitment mechanisms, r and m. The following lemma simplifies the analysis by telling us that workers can never achieve higher utility by using *both* commitment mechanisms than they can achieve using just one or the other.

Lemma 1. *The solution to self 0's maximization problem (2) - (5) includes the contract $(\gamma/2, \gamma/2, 0)$ or it includes a contract of the form $(w, 0, m)$ with $m > 0$.*

All proofs are in the Appendix.

As we have already discussed, a sophisticated worker can achieve first-best through either a non-elective employer contribution contract or a matching contract. The more interesting case is the naive worker. The following lemma tells us that naives always take a matching contract.

Lemma 2. *A naive worker always prefers some zero-profit matching contract of the form $(w, 0, m)$ to the contract $(\gamma/2, \gamma/2, 0)$.*

What level of m will the worker choose? Recall that naive workers have both an overestimation motivation (because $\hat{\beta} > \beta$) and a commitment motivation (if $\hat{\beta} < 1$) for using m. The condition for the m in the equilibrium contract of the form $(w, 0, m)$ can be expressed as:

$$(7) \quad \frac{s(w^*, 0, m^*|\beta) + m^* \frac{\partial s(w^*, 0, m^*|\beta)}{\partial m}}{1 + m^* \frac{\partial s(w^*, 0, m^*|\beta)}{\partial w}} = \frac{s(w^*, 0, m^*|\hat{\beta}) + (1 - \hat{\beta})(1 + m^*) \frac{\partial s(w^*, 0, m^*|\hat{\beta})}{\partial m}}{(1 + m^*)\hat{\beta} + (1 - \hat{\beta})(1 + m^*) \frac{\partial s(w^*, 0, m^*|\hat{\beta})}{\partial w}}.$$

The RHS of (7) is the worker's marginal rate of substitution between m and w while the LHS is the slope of the zero-profit line. Substituting into (7) for the efficient match ($m^* = m^{FB}$) and rearranging yields the condition for equilibrium m to achieve the first-best,

$$(8) \quad \frac{\beta s(w^*, 0, m^{FB}|\beta) + (1 - \beta) \frac{\partial s(w^*, 0, m^{FB}|\beta)}{\partial m}}{\beta + (1 - \beta) \frac{\partial s(w^*, 0, m^{FB}|\beta)}{\partial w}} = \frac{\beta s(w^*, 0, m^{FB}|\hat{\beta}) + (1 - \hat{\beta}) \frac{\partial s(w^*, 0, m^{FB}|\hat{\beta})}{\partial m}}{\hat{\beta} + (1 - \hat{\beta}) \frac{\partial s(w^*, 0, m^{FB}|\hat{\beta})}{\partial w}}.$$

By inspection, it is easy to see that this condition holds if $\hat{\beta} = \beta$. In equilibrium, sophisticated workers get a first-best commitment contract, as we now state formally.

Proposition 1. *In equilibrium,*

    (1) Sophisticated workers choose either a matching contract with $m^* = m^{FB}$ or a non-elective contribution contract with $r^* = \gamma/2$ and achieve the first best.

    (2) Rational workers choose contracts with $r^* \leq \frac{\gamma}{2}$, $w^* = \gamma - r^*$, and $m^* = 0$.

Sophisticated workers understand their need for commitment and will use either r or m to achieve the first-best outcome, as we have discussed. Rational workers, in contrast, will not choose a matching contract. The reason is that matching inefficiently subsidizes second-period consumption, leading to a costly distortion in rationals' intertemporal consumption choices. Rationals are better off receiving their compensation through the lump-sum payments of w and r. They are indifferent among zero-profit contracts with $r \leq \gamma/2$, since they can simply choose savings to achieve the first-best levels of consumption under any such contract.

Consider now a naive worker. In the behavioral contract theory literature, naive present-biased agents generally do not demand first-best commitment contracts (see, e.g., DellaVigna and Malmendier, 2004; Heidhues and Kőszegi, 2010), so one would not expect (8) to hold in general for naive workers, as the following proposition confirms.

Proposition 2. *Assume $u(c_i)$ takes a CRRA form with coefficient of relative risk aversion equal to θ. Then in equilibrium, naive workers choose a matching contract with $m^* > 0$ such that,*

    (1) If $\theta < 1$ then $m^* > m^{FB}$;

    (2) If $\theta = 1$ then $m^* = m^{FB}$;

    (3) If $\theta > 1$ then $m^* < m^{FB}$.

Naive workers are attracted to matching contracts through a mix of commitment motivation and overestimation motivation, but in equilibrium do not in general receive a first-best matching contract. The elasticity of intertemporal substitution, $1/\theta$, determines whether the equilibrium match over- or under-shoots first-best since it reflects the willingness of workers to tolerate unequal

consumption over time. Workers with a relatively high elasticity ($\theta < 1$) have a lower willingness to accept for tolerating such unequal consumption, and their overestimation of the value of the match results in them choosing a match that overshoots the first-best. Conversely, for relatively low elasticity of intertemporal substitution ($\theta > 1$), the equilibrium entails a lower level of matching and second-period consumption than in the first-best.

Rather surprisingly, with $\theta = 1$, in which $u(c_i) = \ln(c_i)$, (8) holds for all $\hat{\beta} \in [\beta, 1]$, and the overestimation motivation produces the first-best commitment contract even for naive workers who underestimate their need for commitment. With log utility, changes in $\hat{\beta}$ have no effect on the shape of workers' indifference curves over m and w. They merely relabel them, as naive workers anticipate achieving higher utility than they actually do. This counter-intuitive result of first-best commitment for naives, however, is a knife-edge special case. We also show in the proof of Proposition 2 that log utility is not only sufficient for producing the first-best contract for all degrees of naivete, it is also necessary. That is, for every other increasing, concave function $u(\cdot)$, there exists $\hat{\beta} \in (\beta, 1]$ such that $m^* \neq m^{FB}$.

Because naive workers anticipate achieving higher utility, naivete has potential effects on labor supply. Suppose we added to the model a labor-leisure choice in which naive workers optimized based on their overestimation of their future savings behavior. Then employers' retirement savings plans would lead naive workers to supply more labor than sophisticated workers. This is in contrast to the effect of forced savings policies like Social Security, which generally *lower* naive workers' labor supply (Kaplow, 2014). This labor supply effect of employer-sponsored retirement savings plans likely increases naive workers' welfare, given pre-existing distortions in labor supply caused by income taxation.

2.3. **Equilibrium Contracts with Heterogenous Types.** Consider now the case of heterogenous worker types, $(\beta, \hat{\beta}) \in \Theta$, in which firms do not observe workers' types. In this section, we assume that $u(c_i) = \ln(c_i)$. We know from Proposition 2 that with log utility, every $(\beta, \hat{\beta})$ type achieves efficiency when contracting on their own, so this assumption means that any deviation from efficiency we show here must result from the interaction among the types.

We focus on the tractable but still analytically rich case with three types: $\Theta = \{(\beta^r, \hat{\beta}^r),$ $(\beta^n, \hat{\beta}^n), (\beta^s, \hat{\beta}^s)\}$. A fraction κ^r of workers are rational exponential discounters ($\hat{\beta}^r = \beta^r = 1$); a fraction κ^n are naively myopic, with $\beta^n < \hat{\beta}^n \le 1$; and a fraction κ^s are myopic but sophisticated, with $\hat{\beta}^s = \beta^s = \beta^n$. Assume that all three types have positive population shares.

The following definition of a competitive equilibrium follows Rothschild and Stiglitz (1976).

Definition 1. *A competitive equilibrium is a set of contracts C^* such that each contract is weakly preferred by at least one type from among the set of contracts, each contract makes nonnegative profits given the types that prefer it, and there does not exist an alternative contract (w', r', m') such that,*

(1) there exists a worker type that strictly prefers (w', r', m') to all contracts in C^*; and

(2) (w', r', m') would make nonnegative profits if it were chosen by the worker types that strictly prefer it.

For purposes of applying the nonnegative-profits condition, we make the simplifying assumption that, for contracts preferred by multiple types, the fraction of each type among the workers who take the contract is proportional to the type's population share (e.g., if a contract is in the set preferred by naives and rationals, then a fraction $\frac{\kappa^r}{\kappa^r + \kappa^n}$ of workers in that contract are rational, and so forth). The following proposition characterizes the competitive equilibria of the model with heterogenous types.

Proposition 3. *Assume $\Theta = \{(\beta^r, \hat{\beta}^r), (\beta^n, \hat{\beta}^n), (\beta^s, \hat{\beta}^s)\}$. Then competitive equilibria exist and in them:*

(1) Sophisticates choose contracts with $w^* = r^* = \frac{\gamma}{2}$ and $s(w^*, r^*, m^*; \beta^s) = 0$.

(2) There exists a cutoff $\bar{\beta}$ with $\beta^n < \bar{\beta} < 1$ such that,

(a) If $\hat{\beta}^n > \bar{\beta}$, then all naives and rationals pool together in matching contracts with $m^* = \frac{1}{\sigma^{rn}} - 2 < m^{FB}$, where $\sigma^{rn} \equiv \frac{\frac{\kappa^r}{2} + \kappa^n \frac{\beta^n}{1+\beta^n}}{\kappa^r + \kappa^n}$ is the average savings rate in the contract.

13

(b) If $\hat{\beta}^n < \bar{\beta}$, then naives choose contracts with $w^* = r^* = \frac{\gamma}{2}$ and $m^* = 0$, while rationals choose contracts with $r^* \leq \frac{\gamma}{2}$, $w^* = \gamma - r^*$, and $m^* = 0$.

If the naives are sufficiently naive ($\hat{\beta}^n > \bar{\beta}$), then they pool together with the rationals in contracts that offer matching contributions paired with a relatively low wage. They do so because they overestimate the amount of matching contributions they will receive, much like in the homogenous type case. The key difference with heterogenous types is that rational workers drive down the wage of the matching contract due to their relatively high savings rate. In the homogenous type case, naive workers are always paid their marginal product of labor, γ. Here, the *average* total compensation of naives and rationals in the pooled contract equals γ. But because rational workers save more than naive workers and therefore receive a greater amount of matching contributions, the rationals receive compensation greater than γ while naives receive less than γ. The pooling contract thus results in cross-subsidization of rational workers by naive workers.

It is easy to see that a fully naive worker ($\hat{\beta}^n = 1$) will be willing to pool with rationals, since the two types then have the same ex ante preferences over contracts. But even a partially naive worker, who understands that she suffers from some degree of myopia, can be drawn into matching contracts that cross-subsidize rationals. To understand why, first note that zero profits means that, relative to an $m = 0$ contract, any contract with $m > 0$ must offer lower $w + r$ by the amount $m\bar{s}$, where $\bar{s} = \sigma^{rn}w$ is the average actual savings level under the contract. This is part of the cost of matching to workers. The benefit comes from anticipated payments in the amount $ms(w, r, m|\hat{\beta})$, where $s(w, r, m|\hat{\beta})$ is the amount the worker anticipates saving. Crucial to evaluating matching contracts, then, is whether the worker anticipates saving more or less than the average savings in the contract. If $\hat{\beta}^n > \bar{\beta}$, then naive workers anticipate saving more than average and in equilibrium pool with rationals. If $\hat{\beta}^n < \bar{\beta}$ then matching no longer seems like a good deal and naive workers abandon matching.

The cross-subsidization results in naives doing worse in this pooling contract than they would in their best separating contract. The welfare losses for naives occur through two channels. Most directly, the cross-subsidization reduces their total compensation. Less obviously, it leads the level

of the match to be lower than first-best, unlike in the homogenous type case with log utility. The reason is that matching is now more "expensive" in the sense that the wage must be lower for any given level of the match, since rationals will utilize the match at a greater rate than the naives. This interaction between types thus also results in an inefficient allocation of consumption across periods for naives, who consume less in the second period than in the first. And while the redistribution caused by matching makes rationals better off, it does so through an inefficient instrument that also distorts rationals' allocation of consumption across the two periods. In equilibrium rationals consume more in the second period than in the first.

The precise contract terms in any equilibrium pooling contract are those preferred by the naives from among all nonnegative-profit pooling contracts. The reason is that, if not, then that contract could enter, make naives strictly better off, and still make nonnegative profits. Rationals prefer this contract to any nonnegative-profit separating contract because of the cross-subsidy it provides.

In addition to determining the attractiveness of pooling to naives, the average savings rate of rationals and naives also affects the equilibrium matching rate in any equilibrium pooling contract. The average savings rate determines the cost, in terms of reductions in $w + r$, of increasing m. The m^* in any equilibrium pooling contract is therefore a decreasing function of the average savings rate. It is immediate from the statement of the proposition that increasing β^n or κ^r will increase the average savings rate of rationals and naives and therefore decrease m^*:

Corollary 1. *The matching rate in any equilibrium pooling contract is decreasing in β^n and κ^r.*

Sophisticated workers, who have correct beliefs about their future savings, understand that matching is a bad deal and never pool in a matching contract with the rationals. They instead choose a contract in which all of their retirement consumption is financed by non-elective employer contributions. They do so because they understand their need for commitment, and non-elective contributions provide that commitment in a non-redistributive way.

2.4. **Matching Caps.** Thus far we have assumed that firms offer simple linear matching contracts. In reality, retirement plans that include matching always also include a cap on the amount of employee contributions that will be matched, typically in the form of some percentage of the

wage. In this section we expand the contract space to allow for a cap on the matched savings, $c \geq 0$. Otherwise, we maintain all of the assumptions from the heterogeneous type model above. Formally, represent a contract as a 4-tuple (w, r, m, c), where a worker saving s attains first-period consumption $w - s$ and second period consumption $r + s + m \min\{s, c\}$. In this expanded contract space, nonnegative profits still requires $w + r + m\bar{s} \leq \gamma$, but $\bar{s}$ now denotes the average level of *matched* savings across all workers in the contract, i.e., $\min\{s, c\}$. We now have the following result.

Proposition 4. *Assume* $\Theta = \{(\beta^r, \hat{\beta}^r), (\beta^n, \hat{\beta}^n), (\beta^s, \hat{\beta}^s)\}$. *Competitive equilibria exist and in them:*

(1) Sophisticates separate into contracts that deliver consumption of $\gamma/2$ in each period.

(2) All naives and rationals pool together in contracts such that,

    (a) Savings are matched at a rate m^* such that $0 < m^* < m^{FB}$, up to a cap $c^* = s(w^*, r^*, m^* | \hat{\beta}^n) > 0$.

    (b) Rationals save at the matching cap.

    (c) Naives anticipate saving at the matching cap but in fact save strictly less than the matching cap.

Allowing matching caps results in naive and rational workers pooling on matching contracts for all levels of naivete. As before, the terms of the equilibrium pooling contract are those most preferred by naives from among all nonnegative-profit pooling contracts. Those terms include a matching cap set at the naives' anticipated savings level. Naive workers therefore anticipate receiving mc in matching contributions, which is the maximum possible matching benefit. Relative to an $m = 0$ contract, any contract with $m > 0$ must offer lower $w + r$ by the amount $m\bar{s}$, where $\bar{s}$ is now the average actual *matched* savings level under the contract. In fact naives will save less than the cap (since $\beta^n < \hat{\beta}^n$), so $\bar{s}$ is less than c and matching always looks like a good deal to naives, whatever their $\hat{\beta}^n$. In fact, they will not be net beneficiaries from matching, and the matching contract will lead to lower experienced utility than the $r = w = \gamma/2$ contract. The

16

ability of employers to cap their matching offers thus strengthens our prediction that naive and rational workers will pool on a matching contract.

Note that our theory implies that many workers will save strictly less than the amount needed to receive the full employer match offered. If all workers received the full match available, then the overestimation mechanism we have identified would break down.

2.5. **Default Contribution Rates.** We turn now to another important feature of the structure of plan contributions: the default rule for employee contributions to the plan. In traditional defined contribution plans like 401(k)s, new hires have to submit paperwork to affirmatively opt in in order to contribute to the plan. In other words, the default contribution rate that applies if workers take no action is zero. Madrian and Shea (2001) studied an employer that adopted instead a positive default by automatically enrolling new hires into contributing 3% of their salary into its retirement plan unless they affirmatively opted out. They found that automatic enrollment dramatically increased the participation rate of new hires, from 37% to 85%. They also found that the majority of participants contributed the default amount when automatically enrolled. Importantly, the strictly positive default contribution rate under automatic enrollment was stickier than the zero default of the original opt-in design. They interpret the power of the default as stemming both from simple inertia of participants and also from some participants interpreting the default contribution rate as implicit advice from the employer about the *right* amount to contribute.

Behavioral economists seized on these findings to advocate that employers adopt automatic enrollment in order to increase savings (e.g., Thaler and Benartzi, 2004; Orszag, Iwry, and Gale, 2006). A group of economists at the Brookings Institution designed and successfully lobbied for the passage of legislative reforms in the Pension Protection Act (PPA) of 2006 to remove regulatory barriers to the adoption of automatic enrollment (see Beshears, Choi, Laibson, Madrian, and Weller, 2010, for an account of the legislative process). Employers have adopted it in droves, with the percentage of Vanguard-administered plans that use automatic enrollment doubling from 15% in 2007 to 36% in 2014 (Vanguard, 2015). The discovery and adoption of automatic enrollment is widely regarded as behavioral economics' greatest public policy success to-date. In this section

we apply our behavioral contract theory approach to analyze employers' incentives in choosing default contribution rates.

Suppose that in addition to offering employer contributions in the contract as in the baseline model above, employers can now also specify a default contribution rate d for their retirement plan. We need to enrich our behavioral type space to account for the documented stickiness of defaults, since myopic and rational workers as modeled above would simply opt out of any default. Suppose that in addition to having a $(\beta, \hat{\beta})$, each worker also has a "default-sensitivity type"— active chooser, procrastinator, or advice taker—denoted by $\theta \in \{a, p, t\}$. Active choosers ($\theta = a$) behave as in our baseline model above. Procrastinators ($\theta = p$) believe in period 0 that they will save according to their $\hat{\beta}$ but in fact in period 1 will save $s = dw$ if $d \in [0, \bar{d}]$. If $d > \bar{d}$, they revert to saving according to their β, since at a sufficiently high level of default savings, even procrastinators will bear the costs of opting out. Finally, advice takers ($\theta = t$) believe in period 0 that they will save according to their $\hat{\beta}$, but in fact in period 1 will save $s = dw$ if $d \in [\underline{d}, \bar{d}]$, with $\underline{d} > 0$, and will save according to their β otherwise. The idea is that both very low and very high defaults are implausible as advice, but within an intermediate range the worker assumes that the default was chosen in an informed way and follows it. This is consistent with the evidence documenting that a higher fraction of workers stay at a low strictly positive default than stay at a default of zero (Madrian and Shea, 2001). These behavioral assumptions could be microfounded, as in Carroll, Choi, Laibson, Madrian, and Metrick (2009), but we adopt this reduced-form approach for simplicity.

Note that we have assumed that at the time of contracting, no worker type believes that the default d in the contract will affect them. There are two motivations for this assumption. First, this can be thought of as a form of naivete by both procrastinators and advice takers. Second, it could stem from inattention—the default rule governing employee contributions to the employer's retirement plan is far down on the list of important factors to consider when choosing among job offers and hence is simply not salient at the time of contracting.

The rest of the model is as above. We focus on the heterogenous type case with two myopia types (rational and naive) crossed with the three default-sensitivity types (active chooser, procrastinator,

and advice-taker): $\Theta = \{(\beta^r, \hat{\beta}^r), (\beta^n, \hat{\beta}^n)\} \times \{a, p, t\}$. Rationals have $\beta^r = \hat{\beta}^r = 1$ and naive myopic workers have $\beta^n < 1$ and $\hat{\beta}^n \in (\beta^n, 1]$. We omit sophisticated myopic workers because they add complexity to the characterization of the equilibrium without adding any insights. Denote the probability of each type by $\kappa^{ij} = \kappa^i \kappa^j$ for $i \in \{r, n\}$ and $j \in \{a, p, t\}$, where we assume for simplicity that myopia type is independent of default-sensitivity type.

The following proposition characterizes the equilibrium of the model, focusing on the case in which there is a relatively large proportion of active choosers.[3]

Proposition 5. *Assume* $\Theta = \{(\beta^r, \hat{\beta}^r), (\beta^n, \hat{\beta}^n)\} \times \{a, p, t\}$. *There exists a* $\underline{\kappa}$ *such that if* $\kappa^a > \underline{\kappa}$ *then competitive equilibria exist and in them all workers pool in contracts such that:*

(1) *Savings are matched at a rate* $m^* > 0$, *up to a finite cap* $c^* = s(w^*, r^*, m^* | \hat{\beta}^n) > 0$.

(2) *The default contribution rate* d^* *is the one that minimizes average worker savings in the contract, given the other terms of the contract, and the default contribution amount,* $d^* w^*$, *is strictly below the contract's matching cap.*

(3) *All workers anticipate saving at the matching cap, however:*

    (a) *Workers who do not make an active choice save at the default.*

    (b) *Naives who make an active choice in fact save strictly less than the matching cap.*

    (c) *Only rationals who make an active choice save at the matching cap.*

As in the case without defaults, naives and rationals pool in the matching contract most preferred by naives. In equilibrium the default is designed to *minimize* worker savings, conditional on the other terms of the contract. The reason is that defaults that reduce savings also reduce the employer's matching contributions, allowing the employer to offer even higher levels of salient forms of compensation. Key to this result, of course, is our assumption that defaults are not salient at the time of contracting, so that their only substantive effect is through relaxing firms' nonnegative-profit constraint. The default contribution amount is always set below the cap on matching, since otherwise it would not reduce matching payments. Note that this implication is in sharp contrast to

---

[3]If there are not enough active choosers, an uninteresting equilibrium arises in which there is no meaningful interaction between the myopia types.

what a paternalistic employer would do, which is to default workers at the savings rate that maximizes their experienced utility, given the other terms in the contract. That would entail setting the default contribution rate exactly at the cap on matched savings (Bernheim, Fradkin, and Popov, 2015). Similarly, there is no incentive for firms to try to make defaults less sticky by forcing all workers to make an active choice, as suggested by Carroll, Choi, Laibson, Madrian, and Metrick (2009). The reason is that doing so would always increase average savings under the contract relative to the optimal default.

To understand what the savings-minimizing default is, note that our assumptions imply that average saving under the equilibrium contract as a function of the default d is given by:

$$\bar{s}(d) = \begin{cases} \left[\kappa^{ra} + \kappa^{rt}\right]s^r + \left[\kappa^{na} + \kappa^{nt}\right]s^n + \left[\kappa^{rp} + \kappa^{np}\right]dw & \text{if} \quad d < \underline{d} \\ \kappa^{ra}s^r + \kappa^{na}s^n + \left[\kappa^p + \kappa^t\right]dw & \text{if} \quad d \in [\underline{d}, \bar{d}] \\ \kappa^r s^r + \kappa^n s^n & \text{if} \quad d > \bar{d}, \end{cases}$$

where s^r and s^n are the savings levels under the contract of rationals and naives, respectively, when they make an active choice. By inspection we can see that the savings-minimizing default for such a contract is either $d = 0$ or $d = \underline{d}$. The firm will offer $d = \underline{d}$ instead of $d = 0$ if the drop in savings from the advice-takers, who now save at $s = \underline{d}w$ instead of actively choosing on their own, is greater than the increase in savings from moving the procrastinators from $s = 0$ to $s = \underline{d}w$, i.e. $\kappa^{rt}s^r + \kappa^{nt}s^n - \kappa^t\underline{d}w > \kappa^p\underline{d}w$.

A key parameter that determines whether $d^* = 0$ or $d^* = \underline{d}$ is naives' present-bias factor β^n, since it affects how much advice-takers will save under $d = 0$ but not under $d = \underline{d}$. This is illustrated in Figure 1. Panel (A) shows the equilibrium match and default regime as a function of β^n. As β^n increases, the average active savings rate increases. As we have discussed, this makes matching more expensive, resulting in lower m^*. Since the lower matching rate applies to more than just the naive active savers, the net effect is to lower overall matching payments, driving up the wage, as shown in panel (C). Finally, recall that equilibrium contract terms are determined by the preferences of the naive workers, who prefer a cap set at their anticipated savings level. As the
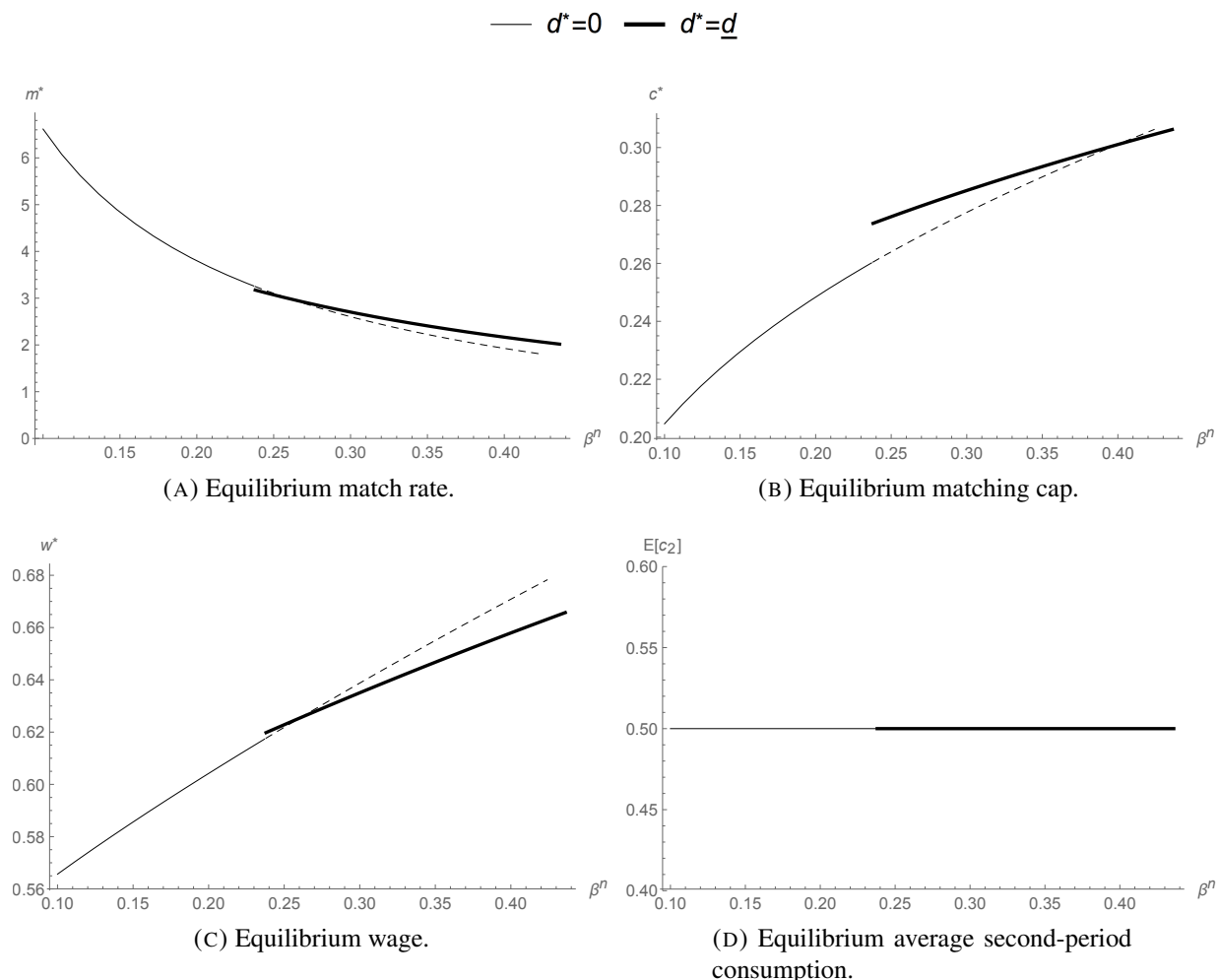
$$\text{—— } d^*=0 \quad \text{—— } d^*=\underline{d}$$

(A) Equilibrium match rate.

(B) Equilibrium matching cap.

(C) Equilibrium wage.

(D) Equilibrium average second-period consumption.

FIGURE 1. Comparative statics on β^n with log utility. The other parameters are $\hat{\beta}^n = 1$, $\kappa^a = 0.6$, $\kappa^t = \kappa^p = 0.2$, $\kappa^r = 0.2$, $\kappa^n = 0.8$, $\underline{d} = 0.12$ and $\gamma = 1$. The thin line denotes $d^* = 0$, thick line denotes $d^* = \underline{d}$, and dashed line denotes outcome if d were set exogenously to 0.

wage increases, the naive anticipates saving more, leading to a higher matching cap, as shown in Panel (B).

By increasing the average savings rate of advice-takers under $d = 0$, but not under $d = \underline{d}$, increasing β^n makes the best nonnegative-profit contract with $d = \underline{d}$ relatively more attractive, resulting in a shift from $d^* = 0$ (indicated by the thin line in the figures) to $d^* = \underline{d}$ (indicated by the thick line). At the switch to automatic enrollment, there is also a discontinuous jump in the matching cap, since that cap will now bind on fewer workers, making it is less "expensive" to raise. The increase in the cap makes average savings go up, making matching more expensive, leading

to a jump down in the match, which in turn leads to a jump up in the wage. We also depict with a dashed line the equilibrium outcome if d were exogenously set at 0, to help illustrate the effect of allowing firms to adopt automatic enrollment.

In our model firms do not use automatic enrollment to paternalistically increase savings, as urged in much of the literature in behavioral economics. This is starkly illustrated in Panel (D), which shows average second-period consumption in equilibrium as a function of β^n. Somewhat counter-intuitively, average second-period consumption is invariant to changes in β^n. More importantly for our purposes, automatic enrollment has no effect on average savings, relative to an opt-in plan. It does have an effect, however, on the *distribution* of savings. When $d^* = \underline{d}$, advice takers save less, and procrastinators save more, relative to if d were exogenously set to 0.

The result that average second-period consumption is invariant to β^n is an artifact of log utility, which generates a constant share of consumption in each period. Log utility is a special case of CRRA utility with the coefficient of relative risk aversion equal to 1. A recent estimation of intertemporal consumption preferences found a coefficient of relative risk aversion of 2.7, a β of 0.35, and an annualized exponential discount factor of 0.97 (Laibson, Maxted, Repetto, and Tobacman, 2015). Accordingly, figure 2 shows the same set of comparative statics using CRRA utility with a coefficient of relative risk aversion equal to 2.7. The results on contract terms in panels (A) - (C) are similar to the results with log utility, but panel (D) shows that equilibrium average second-period consumption is increasing in β^n, which is a more intuitive result. Moreover, it shows that the adoption of automatic enrollment actually *lowers* retirement savings relative to if d were exogenously set to 0—the thick solid line is everywhere below the dashed line.

2.6. **Evidence.** The basic predictions of the model on the structure of plan contributions line up well with key facts on plan design. First, the vast majority of defined contribution plans—about 80%—offer employer matching contributions with a cap on matched savings, and about half provide non-elective employer contributions (PSCA, 2011).

Second, failure to receive the full match offered by the employer is indeed widespread, as implied by our theory. A recent large-scale study found that about 25% of employees did not save
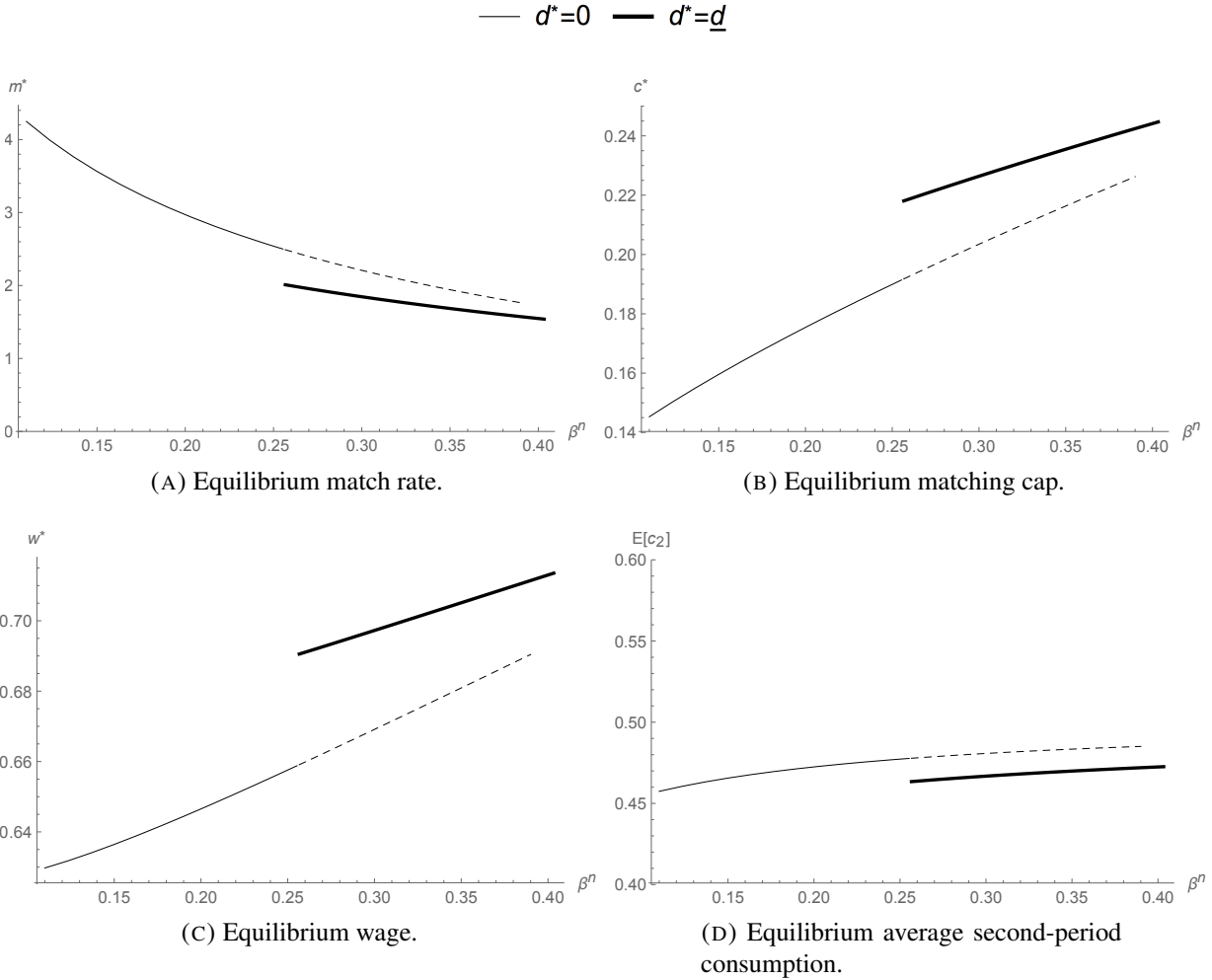
$$d^* = 0 \quad \text{———} \quad d^* = \underline{d}$$

(A) Equilibrium match rate.

(B) Equilibrium matching cap.

(C) Equilibrium wage.

(D) Equilibrium average second-period consumption.

FIGURE 2. Comparative statics on β^n with CRRA utility with coefficient of relative risk aversion equal to 3. The other parameters are $\hat{\beta}^n = 1$, $\kappa^a = 0.8$, $\kappa^t = \kappa^p = 0.1$, $\kappa^r = 0.2$, $\kappa^n = 0.8$, $\underline{d} = 0.12$ and $\gamma = 1$. The thin line denotes $d^* = 0$, thick line denotes $d^* = \underline{d}$, and dashed line denotes outcome if d were set exogenously to 0.

enough to receive their full employer match, foregoing on average \$1,336, or 2.4% of their salary (Financial Engines, 2014).

Third, most employers that have adopted automatic enrollment plans have chosen relatively low default contribution rates. Indeed, about three-quarters of automatic enrollment plans default workers into a 3% initial contribution rate or less (PSCA, 2011), which is the minimum initial contribution rate that qualifies for the safe harbor for such plans created by the PPA.[4] As a result, the adoption of automatic enrollment appears not to have increased overall retirement savings,

---

[4]*See* Internal Revenue Code §401(k)(13)(C)(iii)(I).

contra the hopes of its advocates. While there has been no large-scale evaluation of the effect of the adoption of automatic enrollment on savings rates given the default contribution rates that employers actually choose, Choi, Laibson, Madrian, and Metrick (2004) show that the strength of the competing effects of automatic enrollment on average contribution rates vary from employer to employer and in some cases reduce savings. Furthermore, over the same period that the use of automatic enrollment in plans administered by Vanguard doubled, the average total contribution rate of eligible employees fell from 7.9% to 7.5%, a fall that Vanguard attributes partly to "the growing use of automatic enrollment and the tendency of participants to stick with the default [contribution] rate adopted by the [employer]" (Vanguard, 2015, p.33).

Fourth, automatic enrollment plans almost always offer matching contributions (Beshears, Choi, Laibson, and Madrian, 2010), and the vast majority of those that do set a default contribution rate below the maximum amount of employee contributions that the employer matches (Beshears, Choi, Laibson, and Madrian, 2010, table 4). Butrica and Karamcheva (2012) analyze a sample of 401(k) plans that offer a flat matching rate up to a cap. They find that among automatic enrollment plans, employers' average cap on matching, expressed as a percentage of employee wages, is 5.1%, compared with an average default contribution rate for automatic enrollment plans of only 2.8%. 87% of automatic enrollment plans in their sample use a default contribution rate of 3% or less, whereas 85% of plans use matching caps strictly greater than 3%. Some automatic enrollment plans in their sample default workers into automatically increasing the employee contribution rate over time. They calculate the maximum default contribution rate of each plan as the contribution rate after all such automatic increases. The average maximum default contribution rate, however, is still only 3.4%, well below the average matching cap. These basic stylized facts about the structure of defaults relative to the structure of employer matching contributions are consistent with the positive theory we have developed and inconsistent with a theory in which employers paternalistically set defaults in the interests of workers, given the other terms of the contract.

3. Investment Options

Having developed a positive theory for the rules that govern contributions to employer-sponsored retirement plans, we turn now to the second main feature of retirement plan design: how plan contributions are invested. In a standard self-directed defined contribution plan, the employer offers a menu of investment options (predominantly mutual funds), and employees choose how to invest their plan savings among those options. The median 401(k) plan managed by Vanguard, for example, offers 13 mutual fund options (Tang, Mitchell, Mottola, and Utkus, 2010). Employers typically contract out to service providers for advice on what investment options to offer within the plan as well as for the investment management of those options. Administrative tasks such as recordkeeping and handling contributions to the plan are also typically contracted out.

An overview of these contracting out arrangements is provided by The United States Government Accountability Office (2006, 2012). Many employers contract with a single service provider like Vanguard that offers all of these services in a "bundled" arrangement, but "unbundled" arrangements with multiple service providers are also common. Two main types of fees are charged by these outside service providers. Investment management fees charged on each investment option account for the bulk of total plan fees, while "recordkeeping fees," charged for administering the plan as a whole, account for almost all of the rest. Both types of fees can be charged to either or both the employer and the employee-participants. Investment management fees for mutual funds are typically taken out of each fund's investment return and thereby paid by plan participants. Recordkeeping fees are often paid by employers, at least nominally, but are in some cases paid directly by plan participants. Employers are also commonly given discounts on recordkeeping fees that are funded through the investment management fees charged to plan participants. It is easy to see how this would be possible in the case of a bundled service provider. In unbundled arrangements it is achieved through explicit payments from the investment fund providers for the plan to other plan service providers, an industry practice known as "revenue sharing."

There is evidence that many employee-participants make systematic mistakes in deciding how to allocate their savings across investment options within these plans. Examining a dataset of retirement plans managed by Vanguard, Tang, Mitchell, Mottola, and Utkus (2010) show that while most employers offer an investment menu from which an efficient portfolio can be constructed, participants nonetheless construct inefficient portfolios within these plans that reduce their potential retirement wealth by one-fifth. Similarly, in an experimental setting, Choi, Laibson, and Madrian (2010) show that subjects fail to minimize fees when tasked with allocating a portfolio among four S&P 500 index funds. One underlying source of these mistakes is that many individuals follow a "diversification heuristic," under which, when asked to make several choices at once, they tend to diversify naively. Using a series of experiments as well as observational data, Benartzi and Thaler (2001) document that many individuals follow an extreme form of naive diversification in retirement savings—what they term the "$1/n$ heuristic"—in which, faced with a set of n investment options in their retirement plan, they allocate $1/n$ of their plan contributions to each option. Fisch and Wilkinson-Ryan (2014) show a particularly stark illustration of this phenomenon in a laboratory setting. Subjects were tasked with allocating a retirement portfolio across ten mutual fund options in a hypothetical retirement plan. The menu included two equity index funds that were identical except for fees: one charged 45 basis points while the other only 10 basis points. Using subjects drawn primarily from ivy league undergraduates, they find that 75% of those who invest in the low-fee index fund also invest in the high-fee index fund.

Given this setting in which employers construct the menu of investment options and employees make mistakes in allocating their retirement portfolio among the options, consider then, what incentives do employers have in designing their retirement plan's investment options? We apply our behavioral contract theory approach to analyze this question in the formal model below.

3.1. **Model setup.** To focus on the design of investment options within employer plans, we assume that employers offer compensation contracts composed only of a wage w and a set of investment fund options for the employer's retirement plan, ignoring the employer contributions and default employee contributions analyzed above. Each investment fund is fully characterized by its

investment management fee. Employers choose the number of funds in the plan, N, and the fees charged on each fund, $f^i \geq 0$ for $i = 1, ..., N$, ordered such that $i < j \rightarrow f^i \leq f^j$. We denote a profile of fees charged in a plan by f. We assume that employers in effect receive these fees through discounts on plan administration costs so that the per-worker profits from a contract are given by $\pi(w, f) = \gamma - w + \sum_{i=1}^{N} \bar{s}^i f^i$, where $\bar{s}^i$ is the average savings in the contract that is invested in fund i. The pre-fee gross rate of return of each fund is normalized to 1, and we assume that these returns are certain.[5]

There are three periods which proceed as follows.

- Period 0: Firms make contract offers (w, f) and workers choose among offers.
- Period 1: Workers receive wage w and decide both (1) how much of the wage to save, s; and (2) how to invest their savings.
- Period 2: Retired workers consume their savings.

There are two types of workers, rationals (with probability κ_R) and naives $(1 - \kappa_R)$, each of which have log utility over consumption in each of the two periods, $u(c_1, c_2) = \ln(c_1) + \ln(c_2)$. Rationals are neoclassical maximizers who attend to the investment options that employers offer at the time of contracting. In period 1, rationals decide how to invest their savings among the investment options within the plan and an outside option. We model the outside option as a lower bound on the rate of return on investing within the plan, $r_R \in (0, 1)$, below which rationals would exit the plan and instead save for retirement outside of the plan. This is a type of ex-post participation constraint. Given the tax-deferred treatment of employer-sponsored retirement plans, this outside option would likely entail a similar tax-deferred vehicle such as an IRA. Note that we allow funds within the plan to offer a maximum gross rate of return of 1, whereas $r_R < 1$. Thus, if the employer offers the optimal investment option within the plan, rationals would strictly prefer it to their best outside option. This captures the idea that there are some transaction costs to saving for retirement

---

[5]We could easily allow for uncertain but perfectly correlated returns with no significant changes to the model. We could also allow for imperfectly correlated returns. This would introduce a rational motive for diversification, which would introduce some complications, but the same basic results would hold. In that case, you can think of the model as an analysis of the decision to allocate investments within an asset class, where there are multiple, highly correlated, funds within that class that differ, primarily, in their fee structure.

outside the employer's plan, or alternatively that the tax benefits given to employer-sponsored retirement plans are greater than can be achieved outside such plans, which is true for high-income workers (Bubb, Corrigan, and Warren, 2015).

Naive workers differ from rationals in three ways. First, naives do not attend to the investment options at the time of contracting. Their ex ante utility over contracts in period 0 is simply $u(w, f) = w$. Second, after they take a contract, if naives choose to invest in the plan, they follow the $1/n$ heuristic and divide their savings equally among the fund options in the plan. Third, naives have available an outside option such that they will exit the plan if the average return across funds within the plan, $1 - \bar{f}$, where $\bar{f} \equiv \frac{1}{N} \sum_{i=1}^{N} f^i$, is below a lower bound, $r_N \in (0, 1)$. We assume $r_N < r_R$. This is a reduced-form way to capture the idea that rationals are more sensitive than naives to high fees within the plan and thus more willing to exit in response.

3.2. **Equilibrium contracts.** We use the same Rothschild-Stiglitz equilibrium concept. As a preliminary matter, note that the lowest fee and the (fund-weighted) average fee fully characterize any set of fund options f, given our assumptions about workers' investment behavior. Hence the equilibrium contracts are characterized only up to f^1 and $\bar{f}$, and the model makes no predictions about, for example, the precise number of fund options or other features of the distribution of fees within a plan. We now have our main result.

Proposition 6. *If $r_R > [1 - \frac{1}{2}[\kappa_R(1 - r_R) + (1 - \kappa_R)(1 - r_N)]]^2$, then there exists an equilibrium and in every equilibrium naive and rational workers pool in contracts with $w > \gamma$, $f^1 = 1 - r_R$, and $\bar{f} = 1 - r_N > f^1$, and all workers invest inside the plan.*

The condition in Proposition 6 requires in essence that the wedge between naives' and rationals' reservation price—the most they can be charged within the plan before inducing exit—be sufficiently great. If that is true then in equilibrium both types pool in a contract with a retirement plan characterized by two key features. First is variation in fees among otherwise equivalent investment options. The lowest-fee option charges the level of fees so that the fund's return is equal to rationals' ex post participation constraint, and rationals allocate all of their portfolio to this low-fee option. Other investment options in the plan have higher fees. Note that this price dispersion result

is different from the empirical findings that show large price dispersion in the general retail mutual fund marketplace. There is a surprising degree of price dispersion among S&P 500 index funds, for example (Hortaçsu and Syverson, 2004). Our theory predicts a similar form of price dispersion even among the set of funds chosen by an employer *within a single plan*. Unsurprisingly, the lowest-fee fund in each plan receives disproportionately high investment flows, since rationals allocate all of their portfolio to it whereas naives spread their investment equally across all fund options.

Second, the fund-weighted average fee charged in the plan is relatively high, set so that the rate of return is equal to the lower bound below which naives would be induced to exit the plan. This relatively low lower bound is a reduced form way to capture the stickiness of investing in the plan for naives. That stickiness could come from conventional transaction costs, inertia, or naives' inattentiveness to plan fees.

To see the intuition for why this pooling contract is an equilibrium, first recall that naives simply want a high wage. High plan fees help finance that higher wage since they loosen the employer's zero-profit constraint, which they are held to under perfect competition. Rationals, in contrast, consider fees ex ante, and the efficient fee for rationals in terms of intertemporal consumption choices is zero. They are attracted to the pooling contract with naives, however, because they enjoy cross-subsidization from the higher fees paid by naives, transmitted through a higher wage.

Note that, despite rationals' preference for a fee of zero, in equilibrium rationals pay a fee of $f^1 > 0$. The reason rationals cannot receive a zero-fee investment option in the pooling contract is that that contract would fail free entry: an alternative contract with a slightly higher wage and a slightly higher f^1 could make naives strictly better off without attracting (relatively low fee paying) rationals and still make nonnegative profits. For this deviation not to be possible, we must have $f^1 = 1 - r_R$ so that rationals' ex-post participation constraint binds.

The reason naives cannot keep rationals out of their contract (and avoid cross-subsidizing them) by choosing a contract with uniformly high fees in the plan's investment options is that rationals would still be attracted to the high wage of such a contract and would simply invest using their outside option. In equilibrium the pooling contract offers a sufficiently low fee option to keep

rationals investing in the plan because that subsidizes a higher wage than would be possible if rationals took their outside investment option.

If the condition stated in the proposition fails—consider for example the case in which rationals' and naives' willingness to exit are not significantly different, i.e., $r_R \approx r_N$—then the equilibrium entails separation. With similar willingness to exit, the rationals do not receive much cross-subsidization by pooling with rationals and so prefer to separate into their own plan with efficient (zero) fees.

3.3. **Evidence.** The key predictions of our model line up well with the findings of the largest-scale study to date of the set of investment options offered by employers within 401(k) plans, Ayres and Curtis (2015). The authors analyze a dataset of more than 3,500 plans that includes information on both the set of investment options offered and on the plan-level aggregate portfolio chosen by workers within each plan. They show that while plans generally offer workers a sufficient range of options to achieve an appropriate level of diversification, participants in these plans pay excessive fees, averaging 78 basis points higher than the fees charged on low-cost index funds available in the retail mutual fund marketplace. Even more interestingly, for our purposes, they find that more than half of plans in the sample include so-called "dominated funds," defined as options within the plan menu that have an optimal portfolio weight of less than 1% and that are more than 50 basis points more expensive than funds in the same style either (i) offered within the plan or, if none, (ii) available in the marketplace. Our theory predicts exactly the existence of such dominated funds within employer plans.

4. CONCLUSION

Federal retirement savings policy has long been premised on the notion that, left to their own devices, households will make mistakes in saving for retirement (Kotlikoff, 1987). This paternalistic concern motivates both mandatory savings schemes like Social Security as well as incentive-based policy tools such as tax subsidies for retirement savings that together shape retirement savings in the United States. The special tax subsidies provided for employer-sponsored retirement savings plans amount to an attempt to harness employers to address this policy problem. As a result of these

tax subsidies, each employer designs a microcosm of the broader federal policy regime through the mix of mandatory savings rules, savings incentives, default rules, and investment options they offer workers in their retirement savings plans.

Previous work in economics has considered the problems raised by mandating or subsidizing certain forms of employer benefits such as pensions and health insurance to achieve public policy goals. Summers (1989) argues, for example, that in the presence of wage rigidities, such policies can distort employment levels, in some cases disproportionately harming the very workers the policy seeks to help. Similarly, the predominance of employer-provided health insurance, due in large part to its tax treatment, can cause an inefficient reduction in labor mobility ("job lock") (Gruber, 2000).

We identify a new type of dysfunction caused by attempts to use employers to implement social policy. We show that if workers are subject to behavioral biases that affect retirement savings decisions, then employers have incentives to cater to, rather than correct, those biases. Such biases generally imply a wedge between workers' decision utility at the time of contracting and their experienced utility that is the appropriate criterion for welfare analysis. The equilibrium in the labor market will produce plan designs that maximize the "fictional surplus" measured by workers' ex ante decision utility rather than the true surplus measured by workers' experienced utility. Our analysis thus calls into question the longstanding delegation to employers of the design of the primary tax-advantaged vehicle for retirement savings. If behavioral economists are right that workers make systematic mistakes in saving for retirement, then the labor market will give employers incentives that undermine the field's "public finance" approach to employer plan design. A similar analysis would apply to other forms of employment benefits for which behavioral biases likely play an important role, such as health insurance, but we leave such an analysis for future work.

REFERENCES

AYRES, I., AND Q. CURTIS (2015): "Beyond Diversification: The Pervasive Problem of Excessive Fees and "'Dominated Funds" in 401 (k) Plans.," *Yale Law Journal*, 124(5), 1476–1552.

BAR-GILL, O. (2012): *Seduction by Contract: Law, Economics, and Psychology in Consumer Markets*. Oxford University Press.

BENARTZI, S., AND R. H. THALER (2001): "Naive diversification strategies in defined contribution saving plans," *American economic review*, pp. 79–98.

——— (2007): "Heuristics and biases in retirement savings behavior," *The journal of economic perspectives*, pp. 81–104.

BERNHEIM, B. D., A. FRADKIN, AND I. POPOV (2015): "The welfare economics of default options in 401 (k) plans," *The American Economic Review*, 105(9), 2798–2837.

BESHEARS, J., J. J. CHOI, D. LAIBSON, AND B. C. MADRIAN (2010): "the impact of employer matching on savings plan participation under automatic enrollment," in *Research Findings in the Economics of Aging*, pp. 311–327. University of Chicago Press.

BESHEARS, J., J. J. CHOI, D. LAIBSON, B. C. MADRIAN, AND B. WELLER (2010): "Public policy and saving for retirement: The "autosave" features of the Pension Protection Act of 2006," *Better living through economics*, pp. 274–290.

BUBB, R., P. CORRIGAN, AND P. L. WARREN (2015): "A Behavioral Contract Theory Perspective on Retirement Savings," *Conn. L. Rev.*, 47, 1317.

BUBB, R., AND R. H. PILDES (2014): "How Behavioral Economics Trims Its Sails and Why," *Harvard Law Review*, 127(6), 1637.

BUTRICA, B. A., AND N. S. KARAMCHEVA (2012): "Automatic Enrollment, Employee Compensation, and Retirement Security," *Center for Retirement Research at Boston College Working Paper*, (2012-25).

CARROLL, G. D., J. J. CHOI, D. LAIBSON, B. C. MADRIAN, AND A. METRICK (2009): "Optimal Defaults and Active Decisions," *The Quarterly Journal of Economics*, 124(4), 1639–1674.

CHOI, J. J., D. LAIBSON, AND B. C. MADRIAN (2010): "Why does the law of one price fail? An experiment on index mutual funds," *Review of Financial Studies*, 23(4), 1405–1432.

CHOI, J. J., D. LAIBSON, B. C. MADRIAN, AND A. METRICK (2004): "For Better or Worse: Default Effects and 401 (k) Savings Behavior," *Perspectives on the Economics of Aging*, p. 81.

DELLAVIGNA, S., AND U. MALMENDIER (2004): "Contract design and self-control: Theory and evidence," *The Quarterly Journal of Economics*, pp. 353–402.

——— (2006): "Paying not to go to the gym," *The American Economic Review*, pp. 694–719.

EDLIN, A. S., AND C. SHANNON (1998): "Strict monotonicity in comparative statics," *Journal of Economic Theory*, 81(1), 201–219.

FINANCIAL ENGINES (2014): "Missing out: How much employer 401(k) matching contributions do employees leave on the table?," Discussion paper.

FISCH, J. E., AND T. WILKINSON-RYAN (2014): "Why do retail investors make costly mistakes? An experiment on mutual fund choice," *University of Pennsylvania Law Review*, 162, 605.

GABAIX, X., AND D. LAIBSON (2006): "Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets," *The Quarterly Journal of Economics*, pp. 505–540.

GRUBB, M. D. (2009): "Selling to overconfident consumers," *The American Economic Review*, 99(5), 1770–1807.

GRUBER, J. (2000): "Health insurance and the labor market," *Handbook of health economics*, 1, 645–706.

HEIDHUES, P., AND B. KŐSZEGI (2010): "Exploiting naivete about self-control in the credit market," *The American Economic Review*, pp. 2279–2303.

HORTAÇSU, A., AND C. SYVERSON (2004): "Product Differentiation, Search Costs, and Competition in the Mutual Fund Industry: A Case Study of S&P 500 Index Funds," *The Quarterly Journal of Economics*, pp. 403–456.

KAPLOW, L. (2014): "Government Policy and Labor Supply with Myopic or Targeted Savings Decisions," in *Tax Policy and the Economy, Volume 29*. University of Chicago Press.

KŐSZEGI, B. (2014): "Behavioral contract theory," *Journal of Economic Literature*, 52(4), 1075–1118.

KOTLIKOFF, L. J. (1987): "Justifying public provision of social security," *Journal of Policy Analysis and Management*, 6(4), 674–696.

LAIBSON, D., P. MAXTED, A. REPETTO, AND J. TOBACMAN (2015): "Estimating Discount Functions with Consumption Choices over the Lifecycle," *Unpublished working paper*.

MADRIAN, B. C., AND D. F. SHEA (2001): "The Power of Suggestion: Inertia in 401 (k) Participation and Savings Behavior," *The Quarterly Journal of Economics*.

O'DONOGHUE, T., AND M. RABIN (2001): "Choice and procrastination," *Quarterly Journal of Economics*, pp. 121–160.

ORSZAG, P. R., J. M. IWRY, AND W. G. GALE (2006): *Aging Gracefully: Ideas to Improve Retirement Security in America*. Century Foundation.

PSCA (2011): *54th Annual Survey of Profit Sharing and 401(k) Plans*. Plan Sponsor Council of America.

ROTHSCHILD, M., AND J. STIGLITZ (1976): "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information," *The Quarterly Journal of Economics*, 90(4), 629–649.

SUMMERS, L. H. (1989): "Some simple economics of mandated benefits," *The American Economic Review*, 79(2), 177–183.

TANG, N., O. S. MITCHELL, G. R. MOTTOLA, AND S. P. UTKUS (2010): "The efficiency of sponsor and participant portfolio choices in 401 (k) plans," *Journal of Public Economics*, 94(11), 1073–1085.

THALER, R. H., AND S. BENARTZI (2004): "Save more tomorrow™: Using behavioral economics to increase employee saving," *Journal of political Economy*, 112(S1), S164–S187.

THE UNITED STATES GOVERNMENT ACCOUNTABILITY OFFICE (2006): "Changes Needed to Provide 401(k) Plan Participants and the Department of Labor Better Information on Fees," Discussion paper.

——— (2012): "Increased Educational Outreach and Broader Oversight May Help Reduce Plan Fees," Discussion paper.

VANGUARD (2015): "How America Saves 2015," Discussion paper.

APPENDIX

Proof of Lemma 1. Given some contract (w, r, m), the worker's first-order condition for choice of s is given by,

(9) $\quad \beta(1+m)u'((1+m)s(w,r,m|\beta)+r) - u'(w - s(w,r,m|\beta)) \equiv H(w,r,m,s|\beta) = 0.$

This condition will be satisfied whenever $\beta(1+m)u'(r) \geq u'(w)$, where a strict inequality implies $s(w,r,m|\beta) > 0$.

33

Applying the implicit function theorem, this implies that, in the subset of the contract space such that the first-order condition holds, for each $j \in \{w, r, m\}$,

$$(10) \qquad \frac{\partial s(w, r, m | \beta)}{\partial j} = -\frac{\frac{\partial H}{\partial j}}{\frac{\partial H}{\partial s}},$$

where

$$(11) \qquad \begin{aligned} \frac{\partial H}{\partial r} &= \beta(1 + m)u''(s(1 + m) + r) \\ \frac{\partial H}{\partial w} &= -u''(w - s) \\ \frac{\partial H}{\partial s} &= u''(w - s) + \beta(1 + m)^2 u''((1 + m)s + r). \end{aligned}$$

Implicitly define the zero-profit wage as a function of other contract terms, $w(r, m)$, by $w + ms(w, r, m | \beta) + r - \gamma = 0$. For the subset of the zero-profit contract set such that the first-order condition holds, we have by the implicit function theorem,

$$(12) \qquad \begin{aligned} \frac{\partial w(r, m)}{\partial r} &= -\frac{1 + m\frac{\partial s(w, r, m | \beta)}{\partial r}}{1 + m\frac{\partial s(w, r, m | \beta)}{\partial w}} \\ &= -\frac{\frac{\partial H}{\partial s} - m\frac{\partial H}{\partial r}}{\frac{\partial H}{\partial s} - m\frac{\partial H}{\partial w}} \\ &= -\frac{u''(w - s) + \beta(1 + m)^2 u''((1 + m)s + r) - m\beta(1 + m)u''(s(1 + m) + r)}{u''(w - s) + \beta(1 + m)^2 u''((1 + m)s + r) + mu''(w - s)} \\ &= -\frac{1}{1 + m}. \end{aligned}$$

If the first-order condition for savings is *not* satisfied, however, then $s(w, r, m) = 0$, and $\partial w / \partial r$ is simply -1.

To prove the Lemma, suppose, first, that the solution includes a contract in the portion of the zero-profit set in which the first-order-condition for savings is satisfied for a type-β worker, which implies that it is also satisfied for a type-$\hat{\beta}$ worker, i.e., for anticipated savings (since $\hat{\beta} \geq \beta$). Substituting for the wage in self 0's objective function using $w(r, m)$, we have,

$$(13) \qquad V(r, m | \hat{\beta}) \equiv u(w(r, m) - s(w(r, m), r, m | \hat{\beta})) + u(r + (1 + m)s(w(r, m), r, m | \hat{\beta})).$$

34

Taking the partial derivative with respect to r, we have,

$$V_r = [\frac{\partial w(r,m)}{\partial r} - (\frac{\partial w(r,m)}{\partial r}\frac{\partial s}{\partial w} + \frac{\partial s}{\partial r})]u'(w(r,m) - s(w(r,m),r,m|\hat{\beta}))+$$

$$[1 + (1+m)(\frac{\partial w(r,m)}{\partial r}\frac{\partial s}{\partial w} + \frac{\partial s}{\partial r})]u'(r + (1+m)s(w(r,m),r,m|\hat{\beta})) =$$

$$[-\frac{1}{1+m} - (-\frac{1}{1+m}\frac{\partial s}{\partial w} + \frac{\partial s}{\partial r})]u'(w(r,m) - s(w(r,m),r,m|\hat{\beta}))+$$

(14)
$$[1 + (1+m)(-\frac{1}{1+m}\frac{\partial s}{\partial w} + \frac{\partial s}{\partial r})]u'(r + (1+m)s(w(r,m),r,m|\hat{\beta})) =$$

$$[-\frac{1}{1+m} - (-\frac{1}{1+m}\frac{\partial s}{\partial w} + \frac{\partial s}{\partial r})]\hat{\beta}(1+m)u'(r + (1+m)s(w(r,m),r,m|\hat{\beta}))+$$

$$[1 + (1+m)(-\frac{1}{1+m}\frac{\partial s}{\partial w} + \frac{\partial s}{\partial r})]u'(r + (1+m)s(w(r,m),r,m|\hat{\beta})) =$$

$$[1 + (-\frac{\partial s}{\partial w} + (1+m)\frac{\partial s}{\partial r})](1 - \hat{\beta})u'(r + (1+m)s(w(r,m),r,m|\hat{\beta})),$$

where we have used the first-order condition for anticipated savings. Substituting in for the partial derivatives of anticipated savings, the leading term in brackets can be rewritten as,

(15)
$$[1 + ([\frac{-u''(w-s) - \hat{\beta}(1+m)^2 u''(s(1+m)+r)}{u''(w-s) + \hat{\beta}(1+m)^2 u''((1+m)s+r)}] = 0.$$

Hence, holding m fixed, self 0 is indifferent among all zero-profit contracts for which the first-order condition for savings is satisfied, including the one with $r = 0$. So if the solution includes a contract (w, r, m) in which type-β's FOC for savings is satisfied, then it also includes the contract $(w(0,m), 0, m)$, i.e., with the same m but with $r = 0$ and the corresponding zero-profit wage.

Next, consider the case where the solution includes some contract for which the FOC for savings for the type-$\hat{\beta}$ agent is satisfied but the one for the type-β agent is not. Replicating the argument above, but remembering that in this subset of the contract space $\frac{\partial w(r,m)}{\partial r} = -1$,

$$V_r = [\frac{\partial w(r,m)}{\partial r} - (\frac{\partial w(r,m)}{\partial r}\frac{\partial s}{\partial w} + \frac{\partial s}{\partial r})]u'(w(r,m) - s(w(r,m),r,m|\hat{\beta}))+$$

$$[1 + (1+m)(\frac{\partial w(r,m)}{\partial r}\frac{\partial s}{\partial w} + \frac{\partial s}{\partial r})]u'(r + (1+m)s(w(r,m),r,m|\hat{\beta})) =$$

$$[-1 - (-\frac{\partial s}{\partial w} + \frac{\partial s}{\partial r})]u'(w(r,m) - s(w(r,m),r,m|\hat{\beta}))+$$

(16)
$$[1 + (1+m)(-\frac{\partial s}{\partial w} + \frac{\partial s}{\partial r})]u'(r + (1+m)s(w(r,m),r,m|\hat{\beta})) =$$

$$[-1 - (-\frac{\partial s}{\partial w} + \frac{\partial s}{\partial r})]\hat{\beta}(1+m)u'(r + (1+m)s(w(r,m),r,m|\hat{\beta}))+$$

$$[1 + (1+m)(-\frac{\partial s}{\partial w} + \frac{\partial s}{\partial r})]u'(r + (1+m)s(w(r,m),r,m|\hat{\beta})) =$$

$$[1 - \hat{\beta} - \hat{\beta}m + (1+m)(1-\hat{\beta})(\frac{\partial s}{\partial r} - \frac{\partial s}{\partial w})]u'(r + (1+m)s(w(r,m),r,m|\hat{\beta})),$$

35

Substituting in for the partial derivatives of anticipated savings, the leading term in brackets can be rewritten as,

(17) $\qquad 1 - \hat{\beta} - \hat{\beta}m - (1 - \hat{\beta})[\dfrac{mu''(w - s)}{u''(w - s) + \hat{\beta}(1 + m)^2 u''((1 + m)s + r)} + 1] < 0.$

This implies that the supposed optimal contract is dominated by other contracts in the constraint set with a smaller r, which is a contradiction. This implies that there cannot be a contract in the solution to the problem in which the FOC for the worker's anticipated savings holds but the one for actual savings does not hold.

Finally, suppose there is a contract in the solution for which the FOCs for savings and for anticipated savings are not satisfied. In that case, $w = \gamma - r$ and the worker's problem is simply $\max_r u(\gamma - r) + u(r)$, which is maximised at $r = w = \gamma/2$. If a contract of the form $(\gamma/2, \gamma/2, m)$, in which the FOCs for savings and for anticipated savings are not satisfied, is in the optimal set, then so is the contract $(\gamma/2, \gamma/2, 0)$, since it delivers the same utility. $\qquad \square$

Proof of Lemma 2. Consider the zero-profit matching contract that perfectly offsets the naive's myopia, $(\frac{\gamma(1+\beta)}{2}, 0, \frac{1-\beta}{\beta})$. The self 1 facing that contract will choose s to equate her marginal utility of consumption in the two periods, i.e., so that $u'(\frac{\gamma(1+\beta)}{2} - s) = u'(s/\beta)$. The chosen s results in consumption $\gamma/2$ in each period, verifying zero profits. The sophisticate anticipates this consumption sequence under both the $(\frac{\gamma(1+\beta)}{2}, 0, \frac{1-\beta}{\beta})$ contract and the $(\gamma/2, \gamma/2, 0)$ contract and so is indifferent between them. Note that a myopic worker's period-0 anticipated payoff from the $(\gamma/2, \gamma/2, 0)$ contract is independent of $\hat{\beta}$. In contrast, her anticipated payoff from the $(\frac{\gamma(1+\beta)}{2}, 0, \frac{1-\beta}{\beta})$ contract is strictly increasing in $\hat{\beta}$. To show this last point, we can differentiate the period-0 objective function with respect to $\hat{\beta}$, yielding

$$\frac{\partial}{\partial \hat{\beta}}[u(w - s(w, r, m|\hat{\beta})) + u((1 + m)s(w, r, m|\hat{\beta}))]$$

(18) $\qquad = [(1 + m)u'((1 + m)s(w, r, m|\hat{\beta})) - u'(w - s(w, r, m|\hat{\beta}))]\dfrac{\partial s(w, r, m|\hat{\beta})}{\partial \hat{\beta}}$

$$= u'((1 + m)s(w, r, m|\hat{\beta}))(1 + m)(1 - \hat{\beta})\frac{\partial s(w, r, m|\hat{\beta})}{\partial \hat{\beta}} > 0,$$

where the inequality follows from the fact that $\frac{\partial s(w, r, m|\hat{\beta})}{\partial \hat{\beta}} > 0$. This implies that naive workers, who have $\hat{\beta} > \beta$, strictly prefer the matching contract $(\frac{\gamma(1+\beta)}{2}, 0, \frac{1-\beta}{\beta})$ to the $(\gamma/2, \gamma/2, 0)$ contract. $\qquad \square$

Proof of Proposition 1. The problem for workers with $\hat{\beta} = \beta$ simplifies to,

(19) $\qquad \max_{w,r,m} u(w - s(w, r, m|\beta)) + u(r + (1 + m)s(w, r, m|\beta)),$

subject to,

(20) $\qquad w + r + ms(w, r, m|\beta) = \gamma,$

(21) $\qquad s(w, r, m|\beta) = \arg\max_{s \geq 0} u(w - s) + \beta u(r + (1 + m)s),$

By concavity of the utility function, combined with the budget constraint, the worker can achieve utility no higher than that achieved by perfectly smoothing her consumption across the two periods,

36

$2u(\gamma/2)$. This utility is in fact achieved in any zero-profit contract with $r = w = \gamma/2$ and hence all such contracts must be in the equilibrium set, including $(\gamma/2, \gamma/2, 0)$. This is true for both sophisticated myopic workers and for rational workers.

Any zero-profit contract with $r > \gamma/2$ results in greater consumption in the second period than in the first, and therefore lower utility than $2u(\gamma/2)$, and hence cannot be in the equilibrium set.

Consider finally contracts with $r < \gamma/2$. The FOC for savings is,

$$(22) \qquad u'(w - s(w, r, m)) = \beta(1 + m)u'(r + (1 + m)s(w, r, m)),$$

and must be satisfied so long as $u'(w) \le \beta(1 + m)u'(r)$. If the FOC for savings is satisfied, then the only m that results in perfect consumption smoothing is $m^{FB} = \frac{1-\beta}{\beta}$. Perfect consumption smoothing results in $w - s(w, r, m^{FB}) = (1 + m^{FB})s(w, r, m^{FB}) + r$, or $s(w, r, m^{FB}) = \frac{w-r}{2+m^{FB}}$. At $m = m^{FB}$, the FOC for savings is satisfied so long as $w \ge r$. Hence, any contract with $m = m^{FB}$, $w \ge r$, and $w + r + m^{FB}\frac{w-r}{2+m^{FB}} = \gamma$ is also in the equilibrium.

For sophisticated myopic workers, this completes the proof. For rational workers, note that $\beta = 1$ and hence $m^{FB} = 0$.

Proof of Proposition 2. First, note that we can restrict attention to contracts of the form $(w, 0, m)$ to derive the matching rates of all equilibrium contracts. To see this, note that every contract in the equilibrium set for naives must result in the FOC for savings being satisfied, following the same arguments as in the proofs of Lemmas 1 and 2, which will not be repeated here. Furthermore, if an equilibrium contract (w, r, m) results in the FOC for savings being satisfied, then this implies that there exists a contract $(w', 0, m)$ in the equilibrium set, following the same argument as in the proof of Lemma 1.

The problem then simplifies to:

$$(23) \qquad \max_{w, m} u(w - s(w, 0, m|\hat{\beta})) + u((1 + m)s(w, 0, m|\hat{\beta})),$$

subject to,

$$(24) \qquad w + ms(w, 0, m|\beta) = \gamma$$

$$(25) \qquad s(w, 0, m|\hat{\beta}) = \arg\max_{s \ge 0} u(w - s) + \hat{\beta}u((1 + m)s),$$

$$(26) \qquad s(w, 0, m|\beta) = \arg\max_{s \ge 0} u(w - s) + \beta u((1 + m)s).$$

To economize on notation, in what follows we will suppress the contract arguments of the savings functions, denoting them simply as $s(\beta)$ and $s(\hat{\beta})$.

The first-order conditions for the Lagrangian of this problem are given by:

$$(27) \quad u'(w^* - s(\hat{\beta}))[1 - \frac{\partial s(\hat{\beta})}{\partial w}] + u'((1 + m^*)s(\hat{\beta}))(1 + m^*)\frac{\partial s(\hat{\beta})}{\partial w} + \lambda[-1 - m^*\frac{\partial s(\beta)}{\partial w}] = 0,$$

and

$$(28) \quad -u'(w^* - s(\hat{\beta}))\frac{\partial s(\hat{\beta})}{\partial m} + u'((1+m^*)s(\hat{\beta}))[s(\hat{\beta}) + (1+m^*)\frac{\partial s(\hat{\beta})}{\partial m}] + \lambda[-s(\beta) - m^*\frac{\partial s(\beta)}{\partial m}] = 0.$$

The first-order condition for the problem in (25) is given by:

$$(29) \qquad -u'(w - s(\hat{\beta})) + (1 + m)\hat{\beta}u'((1 + m)s(\hat{\beta})) = 0.$$

Together these imply the following condition, which pins down m^*:

(30)
$$\frac{s(\beta) + m^* \frac{\partial s(\beta)}{\partial m}}{1 + m^* \frac{\partial s(\beta)}{\partial w}} = \frac{s(\hat{\beta}) + (1 - \hat{\beta})(1 + m^*)\frac{\partial s(\hat{\beta})}{\partial m}}{(1 + m^*)\hat{\beta} + (1 - \hat{\beta})(1 + m^*)\frac{\partial s(\hat{\beta})}{\partial w}}.$$

The LHS of (30) is the slope of the zero-profit line in (m, w) space, while the RHS is the worker's marginal rate of substitution between m and w.

Substituting $m^{FB} = \frac{1-\beta}{\beta}$ into (30) and multiplying both sides by $\frac{\beta}{\hat{\beta}}$ yields,

(31)
$$\frac{\beta s(\beta) + (1 - \beta)\frac{\partial s(\beta)}{\partial m}}{\beta + (1 - \beta)\frac{\partial s(\beta)}{\partial w}} = \frac{\beta s(\hat{\beta}) + (1 - \hat{\beta})\frac{\partial s(\hat{\beta})}{\partial m}}{\hat{\beta} + (1 - \hat{\beta})\frac{\partial s(\hat{\beta})}{\partial w}}.$$

This condition is satisfied when $\hat{\beta} = \beta$, so we have shown that $m^* = m^{FB}$ if $\hat{\beta} = \beta$. Our proof strategy is now to derive montone comparative statics on m^* with respect to $\hat{\beta}$ to characterize m^* for naive workers.

It will be useful to rewrite the problem as a univariate maximization problem by substituting in for w using the zero-profit constraint. The problem becomes:

(32)
$$\max_m V(m, \hat{\beta}) \equiv u\big(w(m) - s(\hat{\beta})\big) + u\big((1 + m)s(\hat{\beta})\big),$$

where,

(33)
$$w(m) = \gamma - ms(\beta).$$

CRRA utility is given by $u(c) = \frac{1}{1-\theta}c^{1-\theta}$ where $\theta > 0$ and $\theta \neq 1$. We then have:

(34)
$$u'(c) = c^{-\theta}.$$

(35)
$$u''(c) = -\theta c^{-\theta-1}.$$

and,

(36)
$$u'''(c) = \theta(\theta + 1)c^{-\theta-2}.$$

The first-order condition for savings (29) yields,

(37)
$$s(\hat{\beta}) = \frac{w[\hat{\beta}(1 + m)]^{\frac{1}{\theta}}}{1 + m + [\hat{\beta}(1 + m)]^{\frac{1}{\theta}}},$$

which we can usefully rewrite as $s(\hat{\beta}) = w\sigma(m, \hat{\beta})$, where,

(38)
$$\sigma(m, \hat{\beta}) \equiv \frac{[\hat{\beta}(1 + m)]^{\frac{1}{\theta}}}{1 + m + [\hat{\beta}(1 + m)]^{\frac{1}{\theta}}} = \frac{1}{1 + (1 + m)^{1-\frac{1}{\theta}}\hat{\beta}^{-1/\theta}}.$$

We will need the following partial derivatives of $\sigma(m, \hat{\beta})$:

(39)
$$\frac{\partial \sigma(m, \hat{\beta})}{\partial m} = \frac{-1}{[1 + (1 + m)^{1-\frac{1}{\theta}}\hat{\beta}^{-1/\theta}]^2}(1 - \frac{1}{\theta})(1 + m)^{-1/\theta} = \frac{1 - \theta}{\theta}(1 + m)^{-1/\theta}\sigma^2$$

(40)
$$\frac{\partial \sigma(m, \hat{\beta})}{\partial \hat{\beta}} = \frac{-1}{[1 + (1 + m)^{1-\frac{1}{\theta}}\hat{\beta}^{-1/\theta}]^2}(-\frac{1}{\theta})(\hat{\beta})^{-1-1/\theta} = \frac{\sigma^2}{\theta\hat{\beta}^{1+1/\theta}}$$

$$\text{(41)} \qquad \frac{\partial^2 \sigma(m, \hat{\beta})}{\partial \hat{\beta} \partial m} = \frac{2\sigma\sigma_m}{\theta \hat{\beta}^{1+1/\theta}} = \frac{\sigma_m \sigma_{\hat{\beta}}}{\sigma}$$

With this notation $w = \gamma - mw\sigma(m, \beta)$, and so,

$$\text{(42)} \qquad w = \frac{\gamma}{1 + m\sigma(m, \beta)}.$$

Substituting that back to the savings equation yields,

$$\text{(43)} \qquad s(\hat{\beta}) = \frac{\gamma\sigma(m, \hat{\beta})}{1 + m\sigma(m, \beta)}.$$

This savings function implies the consumption functions,

$$c_1(\hat{\beta}) = w(m) - s(\hat{\beta}) = \frac{\gamma(1 - \sigma(m, \hat{\beta}))}{1 + m\sigma(m, \beta)}$$

(44)

$$c_2(\hat{\beta}) = (1 + m)s(\hat{\beta}) = \frac{\gamma(1 + m)\sigma(m, \hat{\beta})}{1 + m\sigma(m, \beta)}.$$

Overall utility is then given by,

$$V(m, \hat{\beta}) = \frac{1}{1 - \theta}\left[\left[\frac{\gamma(1 - \sigma(m, \hat{\beta}))}{1 + m\sigma(m, \beta)}\right]^{1-\theta} + \left[\frac{\gamma(1 + m)\sigma(m, \hat{\beta})}{1 + m\sigma(m, \beta)}\right]^{1-\theta}\right]$$

(45)

$$= \frac{1}{1 - \theta}\left[\frac{\gamma}{1 + m\sigma(m, \beta)}\right]^{1-\theta}\left[(1 - \sigma(m, \hat{\beta}))^{1-\theta} + ((1 + m)\sigma(m, \hat{\beta}))^{1-\theta}\right].$$

We will now show that if $\theta < 1$ then $\frac{\partial m^*}{\partial \hat{\beta}} > 0$ and therefore $m^* > m^{FB}$. To do so, it will prove useful to work with a monotonic transformation of the objective function, $v(m, \hat{\beta}) \equiv log(V(m, \hat{\beta}))$, which represents the same preferences and hence generates the same choice function $m^*(\hat{\beta})$. It will be sufficient to show that $\frac{\partial^2 v(m, \hat{\beta})}{\partial m \partial \hat{\beta}} > 0$ (Edlin and Shannon, 1998).

Taking logs, our objective function is now,

$$\text{(46)} \quad v(m, \hat{\beta}) = X - (1 - \theta)log(1 + m\sigma(m, \beta)) + log\left[(1 - \sigma(m, \hat{\beta}))^{1-\theta} + ((1 + m)\sigma(m, \hat{\beta}))^{1-\theta}\right],$$

where X is a constant that does not depend on m or $\hat{\beta}$.

Begin with the partial derivative with respect to $\hat{\beta}$:

$$\frac{\partial v(m, \hat{\beta})}{\partial \hat{\beta}} = (1 - \theta)\sigma_{\hat{\beta}}\frac{(1 + m)^{1-\theta}\sigma^{-\theta} - (1 - \sigma)^{-\theta}}{(1 + m)^{1-\theta}\sigma^{1-\theta} + (1 - \sigma)^{1-\theta}}$$

$$= (1 - \theta)\sigma_{\hat{\beta}}\left[\frac{\alpha}{\sigma} - \frac{1 - \alpha}{1 - \sigma}\right]$$

(47)

$$= \frac{(1 - \theta)}{\theta \hat{\beta}^{1+1/\theta}}\sigma^2\left[\frac{\alpha}{\sigma} - \frac{1 - \alpha}{1 - \sigma}\right]$$

$$= \frac{(1 - \theta)}{\theta \hat{\beta}^{1+1/\theta}}\frac{\sigma}{1 - \sigma}[\alpha - \sigma],$$

where

$$\alpha \equiv \frac{[(1+m)\sigma]^{1-\theta}}{[(1+m)\sigma]^{1-\theta} + (1-\sigma)^{1-\theta}}$$

$$= \sigma \Big[\frac{(1+m)^{1-\theta}}{\sigma(1+m)^{1-\theta} + (1-\sigma)(\frac{\sigma}{1-\sigma})^{\theta}} \Big].$$

Note that

$$(\frac{\sigma}{1-\sigma})^{\theta} = \hat{\beta}(1+m)^{1-\theta},$$

so that $\alpha = \frac{\sigma}{\sigma+(1-\sigma)\hat{\beta}}$. Substituting for α in (47), we have,

$$\frac{\partial v(m, \hat{\beta})}{\partial \hat{\beta}} = \frac{(1-\theta)}{\theta \hat{\beta}^{1+1/\theta}} \frac{\sigma^2}{1-\sigma} \Big[\frac{1 - \sigma - (1-\sigma)\hat{\beta}}{\sigma + (1-\sigma)\hat{\beta}} \Big]$$

$$= \frac{(1-\theta)}{\theta \hat{\beta}^{1+1/\theta}} \frac{\sigma^2}{1-\sigma} \Big[\frac{1}{\sigma + \frac{\hat{\beta}}{1+\hat{\beta}}} \Big]$$

We want to show that this function is strictly increasing in m for $\theta < 1$. When $\theta < 1$, all three terms are positive. Furthermore, the leading term is independent of m. Thus, when $\theta < 1$ the function increases in m whenever the product of the last two terms increases in m. To show that the product of the last two terms increases in m, it is sufficient to show that an increasing transformation of the last two terms increases in m. Taking the partial derivative with respect to m of the log of that product, we have,

$$\frac{\partial}{\partial m} log\Big(\frac{\sigma^2}{1-\sigma} \Big[\frac{1}{\sigma + \frac{\hat{\beta}}{1+\hat{\beta}}} \Big] \Big) = \Big[\frac{2}{\sigma} + \frac{1}{1-\sigma} - \frac{1}{\sigma + \frac{\hat{\beta}}{1+\hat{\beta}}} \Big] \sigma_m$$

$$= \Big[\frac{2[\sigma + \frac{\hat{\beta}}{1+\hat{\beta}}](1-\sigma) + \sigma(\sigma + \frac{\hat{\beta}}{1+\hat{\beta}}) - \sigma(1-\sigma)}{\sigma(1-\sigma)(\sigma + \frac{\hat{\beta}}{1+\hat{\beta}})} \Big] \sigma_m$$

$$= \Big[\frac{\frac{\hat{\beta}}{1+\hat{\beta}}(2-\sigma) + \sigma}{\sigma(1-\sigma)(\sigma + \frac{\hat{\beta}}{1+\hat{\beta}})} \Big] \frac{1-\theta}{\theta} (1+m)^{-1/\theta} \sigma^2,$$

which is strictly positive whenever $\theta < 1$.

We now want to show that if $\theta > 1$ then $\frac{\partial m^*}{\partial \hat{\beta}} < 0$ and therefore $m^* < m^{FB}$. We again appeal to monotone comparative statics on m. It is sufficient to show that $\frac{\partial^2 v(m, \hat{\beta})}{\partial m \partial \hat{\beta}} < 0$, which follows using the same basic argument above but noting that for $\theta > 1$, the leading constant in (49) is negative.

All that is left to show is that if $\theta = 1$ then $m^* = m^{FB}$. CRRA utility with $\theta = 1$ is given by $u(c_i) = log(c_i)$. With log utility, savings is simply $s(w|\beta) = w\frac{\beta}{1+\beta}$, independent of m. Thus, $\partial s/\partial m = 0$ and $\partial s/\partial w = \frac{\beta}{1+\beta}$. Furthermore, zero-profit wages satisfy $w^*(m) = \gamma - mw\frac{\beta}{1+\beta} = \frac{\gamma}{1+m\frac{\beta}{1+\beta}}$. Plugging these into the condition for preferring first-best matching (31), it becomes

$$\frac{\beta\frac{\beta}{1+\beta}w^*}{\beta + (1-\beta)\frac{\beta}{1+\beta}} = \frac{\beta\frac{\hat{\beta}}{1+\hat{\beta}}w^*}{\hat{\beta} + (1-\hat{\beta})\frac{\hat{\beta}}{1+\hat{\beta}}},$$

which simplifies to $1/2 = 1/2$, independent of both β and $\hat{\beta}$. This shows that having CRRA utility with $\theta = 1$ is sufficient to produce $m^* = m^{FB}$ for all values of $\hat{\beta}$ and completes the proof of the proposition.

We now prove that log utility is also necessary for this result, as we remark in the text below the proposition. To show this, first recall that the condition for $m^* = m^{FB}$ given in (31) is satisfied for all utility functions if $\hat{\beta} = \beta$. Second, note that the LHS of the condition is independent of $\hat{\beta}$, but the RHS includes terms that depend on $\hat{\beta}$. A necessary condition for $m^* = m^{FB}$ for all values of $\hat{\beta}$ is that the derivative of the RHS of (31) equals zero when evaluated at $\hat{\beta} = \beta$. Denote the RHS of (31) by the function $f(\hat{\beta})$,

$$
(52) \qquad f(\hat{\beta}) \equiv \frac{\beta s(\hat{\beta}) + (1 - \hat{\beta})\frac{\partial s(\hat{\beta})}{\partial m}}{\hat{\beta} + (1 - \hat{\beta})\frac{\partial s(\hat{\beta})}{\partial w}}.
$$

Taking this derivative,

(53)

$$
f'(\hat{\beta}) = \frac{[\beta\frac{\partial s}{\partial \hat{\beta}} + \frac{\partial^2 s}{\partial \hat{\beta}\partial m}(1 - \hat{\beta}) - \frac{\partial s}{\partial m}][\hat{\beta} + (1 - \hat{\beta})\frac{\partial s}{\partial w}] - [(1 - \frac{\partial s}{\partial w}) + (1 - \hat{\beta})\frac{\partial^2 s}{\partial w \partial \hat{\beta}}][\beta s + (1 - \hat{\beta})\frac{\partial s}{\partial m}]}{[\hat{\beta} + (1 - \hat{\beta})\frac{\partial s(\hat{\beta})}{\partial w}]^2}.
$$

The important inputs in this calculation are the partial derivatives of the savings function evaluated at $1 + m = 1/\beta$. We derive them below both in general and evaluated at $\hat{\beta} = \beta$:

$$
(54) \qquad \frac{\partial s}{\partial \hat{\beta}} = -\frac{\hat{\beta}u'(s/\beta)}{\beta^2 u''(w - s) + \hat{\beta}u''(s/\beta)} \xrightarrow{\hat{\beta}=\beta} \frac{-u'}{(1 + \beta)u''}.
$$

$$
(55) \qquad \frac{\partial s}{\partial w} = \frac{\beta^2 u''(w - s)}{\beta^2 u''(w - s) + \hat{\beta}u''(s/\beta)} \xrightarrow{\hat{\beta}=\beta} \frac{\beta}{1 + \beta}.
$$

$$
(56) \qquad \frac{\partial s}{\partial m} = -\frac{\hat{\beta}[\beta^2 u'(s/\beta) + s\beta u''(s/\beta)]}{\beta^2 u''(w - s) + \hat{\beta}u''(s/\beta)} \xrightarrow{\hat{\beta}=\beta} -\frac{\beta}{1 + \beta}[s + \beta\frac{u'}{u''}].
$$

The cross-partials are complex. Denote them $X_j = \frac{\partial}{\partial s}(\frac{\partial s}{\partial j}) \cdot \frac{\partial s}{\partial \hat{\beta}}$ and them them unexpressed for now.

$$
(57) \quad \frac{\partial^2 s}{\partial m \partial \hat{\beta}} = -\frac{\beta^2 u''(w - s)[\beta^2 u'(s/\beta) + s\beta u''(s/\beta)]}{[\beta^2 u''(w - s) + \hat{\beta}u''(s/\beta)]^2} + X_m \xrightarrow{\hat{\beta}=\beta} -\frac{\beta}{(1 + \beta)^2}[s + \beta\frac{u'}{u''}] + X_m.
$$

$$
(58) \qquad \frac{\partial^2 s}{\partial w \partial \hat{\beta}} = \frac{-u''(w - s)u''(s/\beta)\beta^2}{[\beta^2 u''(w - s) + \hat{\beta}u''(s/\beta)]^2} + X_w \xrightarrow{\hat{\beta}=\beta} -\frac{1}{(1 + \beta)^2} + X_w.
$$

Subbing these in, the numerator of (53), evaluated at $\hat{\beta} = \beta$, is given by

(59)

$$[\beta\frac{-u'}{(1+\beta)u''} - \frac{\beta}{(1+\beta)^2}[s+\beta\frac{u'}{u''}](1-\beta) + X_m(1-\beta) + \frac{\beta}{1+\beta}[s+\beta\frac{u'}{u''}]][\beta+(1-\beta)\frac{\beta}{1+\beta}] -$$

$$[(1-\frac{\beta}{1+\beta}) - (1-\beta)\frac{1}{(1+\beta)^2} + (1-\beta)X_w][\beta s - (1-\beta)\frac{\beta}{1+\beta}[s+\beta\frac{u'}{u''}]]$$

$$=[\frac{u'}{u''}[\frac{\beta^2(1+\beta)-\beta^2(1-\beta)-\beta(1+\beta)}{(1+\beta)^2}] + s[\frac{\beta(1+\beta)-\beta(1-\beta)}{(1+\beta)^2}]][\frac{2\beta}{1+\beta}] -$$

$$[\frac{2\beta}{(1+\beta)^2} + (1-\beta)X_w][s\frac{2\beta^2}{1+\beta} - \frac{\beta^2(1-\beta)}{1+\beta}(\frac{u'}{u''})] + \frac{2\beta}{1+\beta}X_m(1-\beta)$$

$$=\frac{u'}{u''}[\frac{2\beta^2(2\beta+1)(\beta-1)}{(1+\beta)^3}] + s\frac{4\beta^3}{(1+\beta)^3} - s\frac{4\beta^3}{(1+\beta)^3} + \frac{2\beta^3(1-\beta)}{(1+\beta)^3}(\frac{u'}{u''}) +$$

$$\frac{\beta(1-\beta)}{1+\beta}[2X_m - \beta X_w(2s - (1-\beta)\frac{u'}{u''})]$$

$$=-\frac{u'}{u''}[\frac{2\beta^2(1-\beta)}{(1+\beta)^2}] + \frac{\beta(1-\beta)}{1+\beta}[2X_m - \beta X_w(2s - (1-\beta)\frac{u'}{u''})]$$

The second term in (59) includes X_m and X_w, which we can calculate as,

(60)

$$X_w = \frac{\partial s}{\partial\hat{\beta}}[\frac{\beta^2[-u'''(w-s)(\beta^2 u''(w-s)+\hat{\beta}u''(s/\beta)) - [-\beta^2 u'''(w-s)+u'''(s/\beta)\frac{\hat{\beta}}{\beta}]u''(w-s)]}{[\beta^2 u''(w-s)+\hat{\beta}u''(s/\beta)]^2}]$$

$$\xrightarrow{\hat{\beta}=\beta} \frac{\partial s}{\partial\hat{\beta}}\frac{\beta^2[-\beta^2 u'''u'' + u''u'''\beta^2 - \beta u'''u'' - u'''u'']}{\beta^2(1+\beta)^2[u'']^2}$$

$$=\frac{\partial s}{\partial\hat{\beta}}\frac{-u'''u''(1+\beta)}{(\beta+1)^2 u''} = \frac{\partial s}{\partial\hat{\beta}}\frac{-u'''}{(1+\beta)u''}.$$

(61)

$$X_m = \frac{\partial s}{\partial\hat{\beta}}(-\hat{\beta})\Big[[2\beta u''(s/\beta)+su'''(s/\beta)][\beta^2 u''(w-s)+\hat{\beta}u''(s/\beta)] -$$

$$[-\beta^2 u'''(w-s)+u'''(s/\beta)\frac{\hat{\beta}}{\beta}][\beta^2 u'(s/\beta)+s\beta u''(s/\beta)]\Big]/[\beta^2 u''(w-s)+\hat{\beta}u''(s/\beta)]^2$$

$$\xrightarrow{\hat{\beta}=\beta} \frac{\partial s}{\partial\hat{\beta}}\frac{-\beta^2[(2\beta u''+su''')u''(\beta+1)+u'''(\beta^2-1)(\beta u'su'')]}{\beta^2(1+\beta)^2[u'']^2}$$

$$=\frac{\partial s}{\partial\hat{\beta}}\frac{u'''(1-\beta)(\beta u'+su'')-u''(2\beta u''+su''')}{(\beta+1)[u'']^2}$$

$$=\frac{\partial s}{\partial\hat{\beta}}\frac{\beta(1-\beta)u'u'''-2\beta[u'']^2-\beta su''u'''}{(\beta+1)[u'']^2}$$

$$=\frac{\partial s}{\partial\hat{\beta}}\frac{\beta(1-\beta)u'u'''-2\beta[u'']^2-\beta su''u'''}{(\beta+1)[u'']^2}.$$

So the second term in in (59) is,

(62)
$$2X_m - \beta X_w(2s - (1-\beta)\frac{u'}{u''}) = \frac{\partial s}{\partial \hat{\beta}}[2(\frac{\beta(1-\beta)u'u''' - 2\beta[u'']^2 - \beta s u''u'''}{(\beta+1)[u'']^2}) - \beta(\frac{-u'''}{(1+\beta)u''})(2s - (1-\beta)\frac{u'}{u''})$$

$$= \frac{\partial s}{\partial \hat{\beta}}\frac{\beta}{1+\beta}[\frac{(1-\beta)u'u'''}{[u'']^2} - 4]$$

$$= -\frac{u'\beta}{u''(1+\beta)^2}[\frac{(1-\beta)u'u'''}{[u'']^2} - 4]$$

The numerator of $f'(\beta)$ in (59) is thus,

$$= \frac{-u'}{u''}[\frac{\beta^2(1-\beta)}{(1+\beta)^3}]\left[2(1+\beta) + \frac{(1-\beta)u'u'''}{[u'']^2} - 4\right]$$

(63)
$$= \frac{-u'}{u''}[\frac{\beta^2(1-\beta)}{(1+\beta)^3}]\left[\frac{(1-\beta)u'u'''}{[u'']^2} - 2(1-\beta)\right]$$

$$= \frac{-u'}{u''}[\frac{\beta^2(1-\beta)^2}{(1+\beta)^3}]\left[\frac{u'u'''}{[u'']^2} - 2\right]$$

A necessary condition for $m^* = m^{FB}$ for all values of $\hat{\beta}$ is that $f'(\beta) = 0$, which requires that $\frac{u'u'''}{[u'']^2} = 2$. This differential equation yields the general solution $u(x) = A\ln(Bx + C) + D$ with positive multiplicative constants. Hence log utility is also *necessary* for this result. □

Proof of Proposition 3.

Lemma 3. *In any competitive equilibrium, the sophisticates receive a payoff of $2\ln(\gamma/2)$.*

Proof. The contract $(\gamma/2, \gamma/2, 0)$ produces a payoff of $2\ln(\gamma/2)$ and makes nonnegative profits. Hence if the sophisticates receive a payoff in equilibrium that is strictly below $2\ln(\gamma/2)$, then free entry is violated, contradicting the assumption of a competitive equilibrium. Suppose instead sophisticates receive an equilibrium payoff strictly greater than $2\ln(\gamma/2)$. Let (w', r', m') denote an equilibrium contract taken by sophisticates. Then $\ln(w' - s(w', r', m'|\beta^s)) + \ln(r' + (1 + m')s(w', r', m'|\beta^s)) > 2\ln(\gamma/2)$. By the concavity of the log function, this implies that $w' - s(w', r', m'|\beta^s) + r' + (1 + m')s(w', r', m'|\beta^s) > 2 * (\gamma/2)$ and so $w' + r' + m's(w', r', m'|\beta^s) > \gamma$. Under any contract, sophisticates save weakly less than any other type, so this implies that average compensation of the types who choose the contract (w', r', m') is also strictly greater than γ. This implies that the contract makes negative profits, contradicting our supposition of an equilibrium. □

Lemma 4. *In any competitive equilibrium, in any contract preferred by only rationals, rationals receive a payoff of $2\ln(\gamma/2)$.*

Proof. For brevity we omit the proof since it follows the proof of Lemma 3 above exactly except that average savings in such an equilibrium contract equals the savings of rationals, since we are considering an equilibrium contract preferred only by rationals. □

Lemma 5. *Suppose $\hat{\beta}^n > \frac{\sigma^{rn}}{1-\sigma^{rn}}$. Then in any competitive equilibrium in which any workers receive employer matching contributions, all rationals and naives pool in contracts with $m = \frac{1-2\sigma^{rn}}{\sigma^{rn}}$,*

43

where $\sigma^{rn} \equiv \frac{\kappa^r \frac{1}{2} + \kappa^n \frac{\beta^n}{1+\beta^n}}{\kappa^r + \kappa^n}$, *and all sophisticates separate in contracts of the form* $\left(\frac{\gamma}{2}, \frac{\gamma}{2}, m^s\right)$ *with* $s\left(\frac{\gamma}{2}, \frac{\gamma}{2}, m^s; \beta^s\right) = 0$.

Proof. We will use the term "matching contract" to refer to a contract in which at least some workers who prefer the contract receive matching contributions (i.e., that has $m > 0$ and also $s(w, r, m|\beta) > 0$ for some worker type $(\beta, \hat{\beta})$ that prefers it).

First, there cannot be any equilibrium matching contract that only rationals prefer. To see this, note that, assuming the contract only attracts rationals, rationals' preferred zero-profit contract is the solution to,

$$(64) \qquad \max_{w,r,m} u(w - s(w, r, m|1)) + u((1 + m)s(w, r, m|1) + r),$$

subject to,

$$(65) \qquad w + r + ms(w, r, m|1) = \gamma,$$

$$(66) \qquad s(w, r, m|1) = \arg\max_{s \geq 0} u(w - s) + u(r + (1 + m)s).$$

It is easy to show that the solution has $m^* = 0$, since any strictly positive match would distort rationals' savings choice. Moreover, the profits of a contract with $m = 0$ are independent of the types that prefer it, since compensation under such a contract is not a function of workers' savings choices. Any proposed equilibrium in which there is an equilibrium contract that only rationals prefer and under which rationals receive matching contributions would thus violate free entry, since rationals would strictly prefer a $m = 0$ zero-profit contract such as $(\gamma, 0, 0)$ to such a matching contract.

This implies that in any competitive equilibrium that includes any matching contracts, myopic workers (sophisticated or naive) must be among those who prefer the equilibrium matching contracts.

Second, in such an equilibrium, there must be a matching contract in which rationals pool with some type of myopic workers. Consider a matching contract (w, r, m) in such an equilibrium. As we have shown, some type of myopic worker must prefer such a matching contract. Such a contract must deliver utility to those myopic workers of at least $2\ln(\gamma/2)$. (Proof: Suppose it delivered utility strictly less than $2\ln(\gamma/2)$. Then myopic workers would strictly prefer the zero-profit contract $(\gamma/2, \gamma/2, 0)$, which delivers utility of $2\ln(\gamma/2)$. This implies that free entry is violated, which is a contradiction.) This implies that for the $\hat{\beta}$'s of the myopic workers who prefer the contract, we must have:

$$(67) \qquad u(w - s(w, r, m|\hat{\beta})) + u((1 + m)s(w, r, m|\hat{\beta}) + r) \geq 2\ln\gamma/2.$$

Taking the partial derivative of the LHS of (67) with respect to $\hat{\beta}$ yields,

$$\frac{\partial}{\partial\hat{\beta}}[u(w - s(w, r, m|\hat{\beta})) + u((1 + m)s(w, r, m|\hat{\beta}) + r)]$$

$$(68) \qquad = [(1 + m)u'((1 + m)s(w, r, m|\hat{\beta})) - u'(w - s(w, r, m|\hat{\beta}))]\frac{\partial s(w, r, m|\hat{\beta})}{\partial\hat{\beta}}$$

$$= u'((1 + m)s(w, r, m|\hat{\beta}))(1 + m)(1 - \hat{\beta})\frac{\partial s(w, r, m|\hat{\beta})}{\partial\hat{\beta}} > 0,$$

where the second equality follows from the first-order condition for savings and the inequality follows from the fact that $\frac{\partial s(w,r,m|\hat\beta)}{\partial\hat\beta} > 0$. Since $\hat\beta^r$ is (at least weakly) greater than $\hat\beta^n$ and $\hat\beta^s$, this implies that rationals would receive a payoff at least weakly greater than $2\ln(\gamma/2)$ from such a contract. But by Lemma 4, in any equilibrium contract in which they do not pool with myopes, rationals receive $2\ln(\gamma/2)$, which means that rationals must prefer pooling with myopes in such a matching contract.

Third, in such an equilibrium, sophisticated myopic workers cannot pool with rationals in a matching contract. Suppose there were an equilibrium matching contract (w', r', m') in which sophisticates pool with rationals. To be an equilibrium, such a contract must earn zero profits, so we must have $w' + r' + m'\bar s = \gamma$, where $\bar s$ is the average savings of workers who prefer the contract. Since rationals save strictly more than sophisticated myopic workers and naive myopic workers save the same amount as sophisticated myopic workers, this implies that sophisticated workers' total compensation is $w' + r' + m's(\hat\beta^s) < \gamma$. This implies that they must receive utility of less than $2\ln(\gamma/2)$, which contradicts Lemma 3, so this cannot be an equilibrium.

This implies that in any such equilibrium, there must be a matching contract preferred by rationals and naives and not by sophisticates. We now proceed to characterize the match of such a contract. To begin, suppose that the contract is such that the first-order condition for savings holds for both naives and rationals so that savings for type β is given by,

$$(69) \qquad s(w, r, m|\beta) = \frac{\beta}{1+\beta}w - \frac{r}{(1+m)(1+\beta)}.$$

Average savings of the naive and rational workers that prefer the contract is given by:

$$(70) \qquad \bar s(w, r, m) = \sigma^{rn}w - (1 - \sigma^{rn})\frac{r}{1+m}.$$

Free-entry implies that any equilibrium contract must make zero profits, so we must have $w + r + m\bar s(w, r, m) = \gamma$. Solving for w gives us,

$$(71) \qquad w = \frac{\gamma - r + (1 - \sigma^{rn})\frac{mr}{1+m}}{1 + m\sigma^{rn}} = \frac{\gamma}{1 + m\sigma^{rn}} - \frac{r}{1+m}.$$

Substituting this expression for w into the saving function above, we have,

$$(72) \qquad s(w, r, m; \beta) = \frac{\beta}{1+\beta}[\frac{\gamma}{1 + m\sigma^{rn}}] - \frac{r}{1+m}.$$

The utility of such an equilibrium contract to a worker of type $(\beta, \hat\beta)$ is given by,

$$
\begin{aligned}
V(\frac{\gamma}{1 + m\sigma^{rn}} &- \frac{r}{1+m}, r, m|\hat\beta) = \\
& \ln(\frac{\gamma}{1 + m\sigma^{rn}} - \frac{r}{1+m} - [\frac{\hat\beta}{1+\hat\beta}[\frac{\gamma}{1 + m\sigma^{rn}}] - \frac{r}{1+m}]) \\
& + \ln((1+m)(\frac{\hat\beta}{1+\hat\beta}[\frac{\gamma}{1 + m\sigma^{rn}}] - \frac{r}{1+m}) + r) \\
& = \ln(\frac{1}{1+\hat\beta}) + 2\ln(\gamma) - 2\ln(1 + m\sigma^{rn}) + \ln(1+m) + \ln(\frac{\hat\beta}{1+\hat\beta}).
\end{aligned}
$$

(73)

Note that the value function does not depend on r. It does depend on m. The m that maximizes (73) is given by:

$$(74) \qquad m^* = \frac{1 - 2\sigma^{rn}}{\sigma^{rn}}.$$

Note that this m^* does not depend on $\hat{\beta}$ and hence is preferred by both naives and rationals. This implies that in this subset of the contract space (i.e., contracts for which naives' and rationals' first-order-condition for savings holds), only contracts with $m = m^*$ can be an equilibrium, since both rationals and naives would strictly prefer the zero-profit contract with m^* to any other zero-profit contract in this subset of the contract space.

We now show that there cannot be an equilibrium matching contract for which naives' and rationals' first-order-condition for savings do not hold. First, we do not need to consider contracts for which $s(w, r, m|1) = 0$ since such a contract would not be a matching contract since neither rationals nor naives would save. Hence the only possible case is one in which naives' first-order condition is not satisfied, which implies that $s(w, r, m|\beta^n) = 0$. Consider then the set of contracts for which $s(w, r, m|1) > 0$ and $s(w, r, m|\beta^n) = 0$. Any equilibrium contract must satisfy zero profits, so we must have,

$$(75) \qquad w = \gamma - r - m\frac{\kappa^r s(w, r, m|1)}{\kappa^r + \kappa^n} = \gamma - r - m\frac{\kappa^r\left[\frac{w}{2} - \frac{r}{2(1+m)}\right]}{\kappa^r + \kappa^n}.$$

Solving this for w yields,

$$(76) \qquad w = \frac{\gamma - r + \frac{m\kappa^r r}{2(1+m)(\kappa^r+\kappa^n)}}{1 + \frac{m\kappa^r}{2(\kappa^r+\kappa^n)}}.$$

Substituting this into the utility function of a worker of type $(\beta, \hat{\beta})$ yields,

$$(77) \quad V(w, r, m|\hat{\beta}) = \ln\left(\frac{\gamma - r + \frac{m\kappa^r r}{2(1+m)(\kappa^r+\kappa^n)}}{1 + \frac{m\kappa^r}{2(\kappa^r+\kappa^n)}} - s(w, r, m|\hat{\beta})\right) + \ln((1+m)s(w, r, m|\hat{\beta}) + r).$$

Consider first the case in which $s(w, r, m|\hat{\beta}) > 0$. This implies that we can substitute in for anticipated savings using the first-order condition:

$$
\begin{aligned}
(78) \qquad V(w, r, m|\hat{\beta}) = {} & \ln\left(\frac{1}{1+\hat{\beta}}\frac{\gamma - r + \frac{m\kappa^r r}{2(1+m)(\kappa^r+\kappa^n)}}{1 + \frac{m\kappa^r}{2(\kappa^r+\kappa^n)}} + \frac{r}{(1+m)(1+\hat{\beta})}\right) + \\
& \ln\left((1+m)\left[\frac{\hat{\beta}}{1+\hat{\beta}}\frac{\gamma - r + \frac{m\kappa^r r}{2(1+m)(\kappa^r+\kappa^n)}}{1 + \frac{m\kappa^r}{2(\kappa^r+\kappa^n)}} - \frac{r}{(1+m)(1+\hat{\beta})}\right] + r\right) \\
= {} & \ln\left(\frac{\hat{\beta}(1+m)}{(1+\hat{\beta})^2}\right) + 2\ln\left(\gamma - \frac{m\kappa^n r}{(1+m)(\kappa^r+\kappa^n)}\right) - \ln\left(1 + \frac{m\kappa^r}{2(\kappa^r+\kappa^n)}\right).
\end{aligned}
$$

Note that this function is strictly decreasing in r. This is true independent of $\hat{\beta}$, so for both naives and rationals. This implies that any equilibrium matching contract in this part of the contract space must have the lowest r possible given m and the zero-profit w that goes with that r and m, since otherwise there would be an alternative zero-profit matching contract strictly preferred by both rationals and naives.

In order for $s(w, r, m|\beta^n) = 0$, the derivative of the naive worker's first period utility function with respect to s must be negative,

$$(79) \qquad \frac{\beta^n}{1+\beta^n}\Big[\frac{\gamma - r + \frac{m\kappa^r r}{2(1+m)(\kappa^r+\kappa^n)}}{1 + \frac{m\kappa^r}{2(\kappa^r+\kappa^n)}}\Big] - \frac{r}{(1+m)(1+\beta^n)} \leq 0,$$

where we have again substituted in for the zero-profit w given m and r. The LHS of (79) is decreasing in r at a constant rate and equals 0 for some $r > 0$. Hence any equilibrium matching contract with $s(w, r, m|\beta^n) = 0$ must have r such that

$$(80) \qquad \frac{\beta^n}{1+\beta^n}\Big[\frac{\gamma - r + \frac{m\kappa^r r}{2(1+m)(\kappa^r+\kappa^n)}}{1 + \frac{m\kappa^r}{2(\kappa^r+\kappa^n)}}\Big] - \frac{r}{(1+m)(1+\beta^n)} = 0,$$

but that implies that the first-order condition for naives' savings is satisfied, hence the equilibrium cannot be one with $s(w, r, m|\hat{\beta}^n) > 0$ in which naives' first-order condition for savings is not satisfied.

All that remains is to consider the possibility that $s(w, r, m|\hat{\beta}^n) = 0$ in such a matching contract. Since $s(w, r, m|1) > 0$ and $m > 0$ in such a contract, zero-profits implies that naives anticipate receiving strictly less than γ in total compensation. But that implies that they would prefer the zero-profit contract $(\gamma/2, \gamma/2, 0)$, which means that this proposed equilibrium violates free entry. This implies that the matching contract preferred by rationals and naives must be such that the first-order condition for savings for both rationals and naives is satisfied.

This implies that in any competitive equilibrium in which there is a contract with $m > 0$ preferred by naives and rationals, all such pooling contracts must offer $m = m^*$ given in (74) above. If the pooling contracts in such an equilibrium all offered a different m, then the supposed equilibrium would violate free entry, since both naives and rationals would strictly prefer a zero-profit pooling contract with m^*. Moreover, if a zero-profit pooling contract with m^* were offered, then it would be strictly preferred to any zero-profit pooling contract with a different m and hence no such other pooling contract can be in the equilibrium set.

Fourth, in such an equilibrium, the only contracts preferred by naives and rationals are pooling contracts of this form. The utility received by a worker of type $(\beta, \hat{\beta})$ in such a pooling contract is given by,

$$(81) \qquad \begin{aligned} V\Big(\frac{\gamma}{1+m^*\sigma^{rn}} - \frac{r}{1+m^*}, r, m^*|\hat{\beta}\Big) &= 2\ln\Big(\frac{\gamma}{2(1-\sigma^{rn})}\Big) + \ln\Big(\frac{1-\sigma^{rn}}{\sigma^{rn}}\Big) + \ln\Big(\frac{\hat{\beta}}{(1+\hat{\beta})^2}\Big) \\ &= 2\ln\Big(\frac{\gamma}{2}\Big) + \ln\Big(\frac{\frac{\hat{\beta}}{1+\hat{\beta}}(1 - \frac{\hat{\beta}}{1+\hat{\beta}})}{(1-\sigma^{rn})\sigma^{rn}}\Big). \end{aligned}$$

Note that this is greater than the payoff of $2\ln(\gamma/2)$ if and only if,

$$(82) \qquad \frac{\hat{\beta}}{1+\hat{\beta}}\Big(1 - \frac{\hat{\beta}}{1+\hat{\beta}}\Big) > (1-\sigma^{rn})\sigma^{rn},$$

which is true if and only if,

$$(83) \qquad \hat{\beta} > \frac{\sigma^{rn}}{1-\sigma^{rn}}.$$

Note that $\sigma^{rn} \in (0, 1/2)$ so that $\frac{\sigma^{rn}}{1-\sigma^{rn}} \in (0, 1)$. This condition always holds for rationals, who have $\hat{\beta} = 1$. Since rationals receive $2\ln(\gamma/2)$ in any equilibrium separating contract (Lemma 4), this implies that in any such equilibrium there cannot be a separating contract for rationals. Similarly, the best naives could do in any non-matching contract would be $2\ln(\gamma/2)$, so if $\hat{\beta}^n > \frac{\sigma^{rn}}{1-\sigma^{rn}}$, which we assume in the state of the lemma, then such an equilibrium cannot include a non-matching contract for naives.

Consider finally the possibility of an equilibrium matching contract preferred by naives and not rationals. In such a contract, any type $(\beta, \hat{\beta})$ that anticipated saving a strictly positive amount under the contract would receive utility,

(84)
$$V(w, r, m|\hat{\beta}) = \ln(\frac{1}{1+\hat{\beta}}) + \ln(\frac{\hat{\beta}}{1+\hat{\beta}}) + 2\ln((1+m)w + r).$$

Note that this utility is additively separable in the preference parameters and the contract terms. Note as well that in such a contract both naives and rationals would anticipate saving a strictly positive amount (since $\hat{\beta}^r \geq \hat{\beta}^n > \beta^n$). Together this implies that if naives prefer this contract from among the set of equilibrium contracts, then the rationals do as well. But this contradicts the supposition that the contract is preferred by naives and not rationals.

Fifth, all sophisticates must separate in contracts of the form $(\frac{\gamma}{2}, \frac{\gamma}{2}, m^s)$ with $s(\frac{\gamma}{2}, \frac{\gamma}{2}, m^s|\beta^s) = 0$. We have already shown that sophisticates must separate, since they cannot pool with rationals and all naives pool with rationals. Further, we know from Lemma 3 that they must receive a payoff of $2\ln(\gamma/2)$ in equilibrium. The set of non-matching contracts that achieve this utility is the set of contracts of the form $(\frac{\gamma}{2}, \frac{\gamma}{2}, m^s)$ with $s(\frac{\gamma}{2}, \frac{\gamma}{2}, m^s|\beta^s) = 0$. All that remains is to show that they cannot achieve this utility in a separating matching contract. To achieve this utility for sophisticates, a separating matching contract must offer $m = m^{FB} = \frac{1-\beta^s}{\beta^s}$.

To show this, it will be sufficient to show that rationals would prefer such a contract to their m^* pooling contract with naives, so that this separating matching contract for sophisticates cannot be in the equilibrium. The w and r of such a contract would have to be such that when taken only by sophisticates, it earns zero-profits. (Otherwise, sophisticates could not achieve utility of $2\ln(\gamma/2)$ with it.) The savings of sophisticates under this contract are given by,

(85)
$$s(w, r, m^{FB}|\beta^s) = \frac{\beta^s}{1+\beta^s}(w - r).$$

Zero-profits thus requires,

(86)
$$w + r + \frac{1-\beta^s}{1+\beta^s}(w - r) = \gamma,$$

which gives us,

(87)
$$w = \frac{\gamma}{2}(1 + \beta^s) - r\beta^s,$$

(88)
$$s(\frac{\gamma}{2}(1 + \beta^s) - r\beta^s, r, m^{FB}|\beta^s) = \beta^s(\frac{\gamma}{2} - r).$$

and,

(89)
$$s(\frac{\gamma}{2}(1 + \beta^s) - r\beta, r, m^{FB}|1) = \frac{\gamma}{4}(1 + \beta^s) - r\beta^s.$$

Rational workers receive the following utility from such a contract:

(90)

$$V(\frac{\gamma}{2}(1+\beta^s) - r\beta^s, r, m^{FB}|1) = \ln(\frac{\gamma}{2}(1+\beta^s) - r\beta^s - \frac{\gamma}{4}(1+\beta^s) + r\beta^s) + \ln(\frac{1}{\beta^s}(\frac{\gamma}{4}(1+\beta^s) - r\beta^s) + r)$$

$$= 2\ln(\frac{\gamma}{4}) + 2\ln(1+\beta^s) - \ln(\beta^s)$$

$$= 2\ln(\frac{\gamma}{4}) + \ln(\frac{1}{\frac{\beta^s}{1+\beta^s}(1 - \frac{\beta^s}{1+\beta^s})}).$$

Rationals workers' utility from the pooling contract with naives is given by,

(91) $$V(\frac{\gamma}{1+m^*\sigma^{rn}} - \frac{r}{1+m^*}, r, m^*|1) = 2\ln(\frac{\gamma}{4}) + \ln(\frac{1}{\sigma^{rn}(1-\sigma^{rn})}).$$

Note that $\sigma^{rn} > \frac{\beta^s}{1+\beta^s}$, which implies that rationals' utility from the supposed separating matching contract for sophisticates is higher than their utility from their pooling contract with naives, which is a contradiction with the supposed equilibrium.

□

Lemma 6. *In any competitive equilibrium in which no workers receive matching contributions, rationals choose contracts in the set $\{(\gamma - r^*, r^*, 0) : r^* \leq \gamma/2\}$, sophisticates choose the contract $(\frac{\gamma}{2}, \frac{\gamma}{2}, 0)$, and if $\hat{\beta}^n < 1$ then naives choose the contract $(\frac{\gamma}{2}, \frac{\gamma}{2}, 0)$.*

Proof. First, consider the rational workers. In any competitive equilibrium in which no workers receive matching contributions, rational workers must choose contracts that make zero profits. Obviously in equilibrium they cannot chose contracts that make negative profits. If they chose a contract that made strictly positive profits, then free entry would be violated. The equilibrium contracts for rational workers in such an equilibrium must be in the set of zero-profit non-matching contracts that maximize their utility, since otherwise free entry would be violated. They thus must be the solution to the following constrained optimization problem,

(92) $$\max_{w,r,m} \ln(w - s(w, r, m|1)) + \ln(r + (1+m)s(w, r, m|1)),$$

subject to,

(93) $$w + r = \gamma,$$

and

(94) $$ms(w, r, m|1) = 0,$$

where the second constraint is that rationals not receive matching contributions under the contract. Note that this is the same problem as the optimization problem in the homogenous type case for rationals, given in (2) - (5) but with the added constraint that $ms(w, r, m|1) = 0$. Proposition 1 tells us that the solution without that added constraint is $\{(\gamma/2 - r^*, r^*, 0) : r^* \leq \gamma/2\}$. Note that this solution satisfies $ms(w, r, m|1) = 0$ and so is also the solution to the more constrained problem above.

Second, consider sophisticated myopic workers. Our analysis is similar to our analysis of rational workers. Sophisticated workers must choose contracts that solve the problem in (2) - (5) but with the added constraint that $ms(w, r, m|\beta^s) = 0$. The solution without that constraint, given in

49

Proposition 1, includes a set of contracts that satisfy $ms(w, r, m|\beta^s) = 0$, namely $(\frac{\gamma}{2}, \frac{\gamma}{2}, m^s)$ with $s(\frac{\gamma}{2}, \frac{\gamma}{2}, m^s|\beta^s) = 0$.

However, suppose that $m^s > 0$. Then $s(\frac{\gamma}{2}, \frac{\gamma}{2}, m^s|1) > 0$ and rationals would prefer this contract to any non-matching contract. But that is a contradiction with our supposition of an equilibrium in which no workers receive matching contributions, so sophisticates must take the contract $(\frac{\gamma}{2}, \frac{\gamma}{2}, 0)$.

Finally consider naive myopic workers. In such an equilibrium, naive workers must similarly choose contracts that are in the solution to,

$$(95) \qquad \max_{w,r,m} \ln(w - s(w, r, m; \hat{\beta}^n)) + \ln(r + (1 + m)s(w, r, m; \hat{\beta}^n)),$$

subject to,

$$(96) \qquad w + r = \gamma,$$

and

$$(97) \qquad ms(w, r, m|\beta^n) = 0.$$

We solve this problem by dividing the constraint set in two. First consider the subset of the constraint set for which $m = 0$. Zero-profits gives us $w = \gamma - r$. Anticipated savings is then given by the first-order condition, $s(w, r, m|\hat{\beta}^n) = \frac{\hat{\beta}^n}{1+\hat{\beta}^n}[\gamma - 2r]$. The optimization problem becomes,

$$(98) \qquad \max_r \ln(\gamma - r - \frac{\hat{\beta}^n}{1 + \hat{\beta}^n}[\gamma - 2r]) + \ln(r + \frac{\hat{\beta}^n}{1 + \hat{\beta}^n}[\gamma - 2r]).$$

The first-order condition for this problem is,

$$(99) \qquad \frac{-1 + 2\frac{\hat{\beta}^n}{1+\hat{\beta}^n}}{\gamma - r - \frac{\hat{\beta}^n}{1+\hat{\beta}^n}[\gamma - 2r]} + \frac{1 - 2\frac{\hat{\beta}^n}{1+\hat{\beta}^n}}{r + \frac{\hat{\beta}^n}{1+\hat{\beta}^n}[\gamma - 2r]} = 0,$$

which simplifies to $r = \frac{\gamma}{2}$. So in this subset of the contract space, the solution is $(\gamma/2, \gamma/2, 0)$.

Consider now the remaining subset of the constraint set, which has $m > 0$. In order to satisfy $ms(w, r, m|\beta^n) = 0$, these must have $s(w, r, m|\beta^n) = 0$. Consider the contracts in this set such that $s(w, r, m|1) > 0$. Zero-profits gives us $w = \gamma - r$. But at such a (w, r, m), rationals would strictly prefer it to a non-matching contract for rationals, since they would receive greater than γ in total compensation. That would result in this contract making negative profits and also contradict our supposition that it is a non-matching contract. In order to keep rationals out of the contract, the contract would have to offer a $w < \gamma - r$, but then it would produce lower utility for naives than the $(\gamma/2, \gamma/2, 0)$ contract above. So to be in an equilibrium, contracts in this subset must be such that $s(w, r, m|1) = 0$, which requires that $\frac{w}{2} - \frac{r}{2(1+m)} \le 0$. This simplifies to $m \le \frac{2r-\gamma}{\gamma - r}$. If $r < \frac{\gamma}{2}$, this implies that $m < 0$, which is not feasible. If $r = \frac{\gamma}{2}$, then this implies that $m = 0$, so we cannot have an equilibrium contract for naives with $m > 0$. So in such an equilibrium, naives prefer the contract $(\gamma/2, \gamma/2, 0)$.

$\square$

Lemma 7. *If $\hat{\beta}^n > \frac{\sigma^{rn}}{1-\sigma^{rn}}$ then in any competitive equilibrium there are workers who receive matching contributions and there exists a competitive equilibrium.*

Proof. In any competitive equilibrium in which no workers receive matching contributions, the highest utility any type can achieve is $2\ln(\gamma/2)$. (Proof: To satisfy nonnegative profits, every contract in such an equilibrium must have $w + r \le \gamma$. By concavity of the per-period utility function,

for any level of total compensation, worker's utility is maximized by equalizing consumption in the two periods. Hence the highest utility they can achieve is to consume $\gamma/2$ in each period.) For this to be an equilibrium, it must satisfy free entry. Consider the matching contract $(\frac{\gamma}{2(1-\sigma^{rn})}, 0, \frac{1-2\sigma^{rn}}{\sigma^{rn}})$, where $\sigma^{rn} \equiv \frac{\frac{\kappa^r}{2}+\kappa^n\frac{\beta}{1+\beta}}{\kappa^r+\kappa^n}$. This is one of the zero-profit pooling contracts taken by naives and rationals in an equilibrium that includes matching contracts, characterized above. The utility received by a worker of type $(\beta, \hat{\beta})$ in such a contract is given by,

$$(100) \qquad V(\frac{\gamma}{2(1-\sigma^{rn})}, 0, \frac{1-2\sigma^{rn}}{\sigma^{rn}}|\hat{\beta}) = 2\ln(\frac{\gamma}{2}) + \ln(\frac{\frac{\hat{\beta}}{1+\hat{\beta}}(1-\frac{\hat{\beta}}{1+\hat{\beta}})}{(1-\sigma^{rn})\sigma^{rn}}).$$

Note that this is greater than the payoff of $2\ln(\gamma/2)$ if and only if $\hat{\beta} > \frac{\sigma^{rn}}{1-\sigma^{rn}}$, which is true for rationals and in the statement of the lemma we have assumed is true for naive myopic workers. So there cannot be a competitive equilibrium in which no workers receive matching contributions. Hence in any competitive equilibrium some workers receive matching contributions.

It remains to show existence of a competitive equilibrium. We do so by giving an example. Consider the set of contracts $\{(\gamma/2, \gamma/2, 0), (\frac{\gamma}{2(1-\sigma^{rn})}, 0, \frac{1-2\sigma^{rn}}{\sigma^{rn}})\}$. Among the contracts in this set, the first is preferred by sophisticates while the second is preferred by rationals and naives. To show that this set satisfies free entry, note first that any contract that would make sophisticates strictly better off would have to make negative profits, following the argument in the proof of Lemma 3. Second, note that similarly any alternative contract that made only rationals strictly better off would also have to make negative profits, following the argument in the proof of Lemma 4. Third, note that any contract that attracts the naives will also attract rationals, to see this note that the utility at type $(\beta, \hat{\beta})$ anticipates receiving from a contract with terms (w, r, m) with non-binding r is given by

$$(101) \qquad \begin{aligned} V(w, r, m|\hat{\beta}) &= \ln(\frac{w}{1+\hat{\beta}}) + \ln((1+m)\frac{\hat{\beta}w}{1+\hat{\beta}}) \\ &= 2\ln(w) + \ln(1+m) + \ln(\hat{\beta}/(1+\hat{\beta})^2), \end{aligned}$$

so if $V(w', r', m'|\hat{\beta}) > V(w, r, m|\hat{\beta})$ for some $\hat{\beta}$ then it is also true for any other $\hat{\beta}$, so it is impossible for a entering contract to attract the rationals without also attracting the naives.[6]

The only remaining type of alternative contract that could lead to a violation of free entry is a contract that makes both rational workers and naive myopic workers strictly better off. We have already shown in the proof of Lemma 4 that there is no zero-profit matching contract that makes both rationals and naives strictly better off than the pooling matching contract in this proposed equilibrium. Furthermore, a non-matching contract that made nonnegative profits could deliver utility of at most $2\ln(\gamma/2)$ for rationals and naives, and we have already shown that under our assumption $\hat{\beta}^n > \frac{\sigma^{rn}}{1-\sigma^{rn}}$, both rationals and naives achieve utility greater than $2\ln(\gamma/2)$ in their equilibrium pooling contract. $\qquad\square$

Lemma 8. *If $\hat{\beta}^n < \frac{\sigma^{rn}}{1-\sigma^{rn}}$ then in any competitive equilibrium there are no workers who receive matching contributions and there exists a competitive equilibrium.*

---

[6]A contract with a binding r can never deliver more the $2\ln(\gamma/2)$, so will never attract the naives.

Proof. Lemma 5 and its proof show that the payoff to rationals and naives in any competitive equilibrium in which any workers receive matching contributions is given by,

$$(102) \qquad V\left(\frac{\gamma}{2(1-\sigma^{rn})}, 0, \frac{1-2\sigma^{rn}}{\sigma^{rn}} | \hat{\beta}\right) = 2\ln\left(\frac{\gamma}{2}\right) + \ln\left(\frac{\frac{\hat{\beta}}{1+\hat{\beta}}(1-\frac{\hat{\beta}}{1+\hat{\beta}})}{(1-\sigma^{rn})\sigma^{rn}}\right).$$

$\hat{\beta}^n < \frac{\sigma^{rn}}{1-\sigma^{rn}}$ implies that this payoff for naive myopic workers is strictly less than $2\ln(\gamma/2)$, which naive workers would receive from the zero-profit contract $(\gamma/2, \gamma/2, 0)$. This implies that there cannot be such an equilibrium, and hence that in any competitive equilibrium there are no workers who receive matching contributions.

To show existence, consider the set $\{(\gamma/2, \gamma/2, 0)\}$. This would earn zero profits and produce utility of $2\ln(\gamma/2)$ for every worker type. To show that this set satisfies free entry, note first that any contract that would make sophisticates strictly better off would have to make negative profits, following the argument in the proof of Lemma 3. Second, note that similarly any alternative contract that made only rationals strictly better off would also have to make negative profits, following the argument in the proof of Lemma 4. Third, our analysis of equations (67) and (68) above shows that any alternative contract that makes naive myopic workers strictly better off must also make rational workers strictly better off. Finally, in the proof of Lemma 5 we show that the highest payoff to rationals and naives that can be achieved in a zero-profit contract preferred by rationals and naives is given by (102) above. $\hat{\beta}^n < \frac{\sigma^{rn}}{1-\sigma^{rn}}$ implies that this payoff for naive myopic workers is strictly less than $2\ln(\gamma/2)$, which naive workers would receive from the equilibrium contract $(\gamma/2, \gamma/2, 0)$. Hence our proposed equilibrium satisfies free entry.

□

Lemmas 5, 6, 7, and 8 together imply Proposition 3. □

Proof of Proposition 4. The proof of this proposition follows that of Proposition 3. The proofs of Lemmas 3, 4, and 6 regarding non-matching contracts follow through nearly identically and will not be repeated here. The key difference introduced by matching caps concerns how matching contracts work. We will first characterize the naive's preferred nonnegative-profit contracts, conditional on being pooled with the rationals. We will then argue that a set of contracts that includes one (or more) from this set, along with the contract $(\gamma/2, \gamma/2, 0, 0)$ is an equilibrium. Finally, we show that any equilibrium must include contracts from this set and a contract preferred only by the sophisticates that delivers consumption $c_1 = c_2 = \gamma/2$, completing the proof.

Lemma 9. *Consider the set of all contracts that make nonnegative profits when chosen by only the naives and rationals. The naives' preferred contracts in that set are matching contracts with a matching cap set at the naive's anticipated savings level, in which rationals will save exactly to the cap but naives will save strictly less than the cap. They use $m < m^{FB}$ and deliver to both naives and rationals identical utility exceeding $2\ln(\gamma/2)$.*

Proof. For this proof, to save on notation, normalize so that $\kappa^r + \kappa^n = 1$.

Note that the naives' preferred contracts in this set must make zero profits since otherwise the contract's wage could be increased to make naives strictly better off while still making nonnegative profits. Hence we can restrict attention to contracts that would make zero profits when chosen by only naives and rationals.

Assume for now that any contract in the constraint set that delivers the highest utility to naives—which we will refer to as an "optimal contract" for short—is a contract in which naive workers anticipate receiving matching contributions (a "matching contract"), so that $m\min(s(w,r,m,c|\hat{\beta}^n),c) > 0$. This requires that $w\hat{\beta}^n > r/(1+m)$ and $m > 0$. We will show later that naives' payoff from the best matching contacts exceeds $2\ln(\gamma/2)$. Since $2\ln(\gamma/2)$ is the highest payoff possible in a non-matching contract (see argument in the proof of Lemma 6), this implies that any optimal contract must be a matching contract.

First, characterize savings behavior of a type-β worker for a given contract, $s(w,r,m,c|\beta)$. If c is small enough, the worker will save above the cap. In that case, savings satisfies $u'(w - s) = \beta u'(s + mc + r)$, which implies savings of $\frac{\beta w - mc - r}{1+\beta}$, which is, in fact, above c whenever $c < \frac{w\beta - r}{1+\beta+m}$. At the other extreme, c is large enough it will not bind. In that case, savings satisfies $u'(w - s) = (1 + m)\beta u'(s(1 + m) + r)$, which implies savings of $\max\{\frac{\beta w - r/(1+m)}{1+\beta}, 0\}$, which must be below c. Between these cutoffs, the agent will simply save to the cap. In summary,

$$(103) \qquad s(w,r,m,c|\beta) = \begin{cases} \frac{\beta w - mc - r}{1+\beta} & \text{, if } c < \frac{w\beta - r}{1+\beta+m} \\ c & \text{, if } \frac{w\beta - r}{1+\beta+m} \le c \le \frac{\beta w - r/(1+m)}{1+\beta} \\ \max\{\frac{\beta w - r/(1+m)}{1+\beta}, 0\} & \text{, if } \frac{\beta w - r/(1+m)}{1+\beta} < c. \end{cases}$$

Denote these cutoff matching caps, $\underline{c}(\beta) = \frac{w\beta - r}{1+\beta+m}$ and $\overline{c}(\beta) = \frac{\beta w - r/(1+m)}{1+\beta}$. It is useful to note that $\underline{c}(\beta) < \overline{c}(\beta)$ unless $m = 0$, both increase in β, and $\underline{c}(\hat{\beta}) > \overline{c}(\beta)$ iff $\hat{\beta} > (1+m)\beta$.

The constrained optimization problem is:

$$(104) \qquad \max_{w,r,m,c} V(w,r,m,c) = \ln(w - s(w,r,m,c|\hat{\beta}^n))$$
$$+ \ln(s(w,r,m,c|\hat{\beta}^n) + m\min(s(w,r,m,c|\hat{\beta}^n),c) + r),$$

subject to,

$$(105) \qquad w = \gamma - \bar{s}(w,r,m,c)m - r,$$

where $\bar{s}(w,r,m,c) = \kappa^r \min\{c, s(w,r,m,c|1)\} + \kappa^n \min\{c, s(w,r,m,c|\beta^n)\}$ is the average matched savings.

Implicitly define the zero-profit wage as a function of the other contract parameters,

$$(106) \qquad w^*(r,m,c) = \gamma - \bar{s}(w^*(r,m,c),r,m,c)m - r$$

The optimization problem can now be rewritten as,

$$(107) \qquad \max_{r,m,c} V(r,m,c) = \ln(w^* - s(w^*,r,m,c|\hat{\beta}^n))$$
$$+ \ln(s(w^*,r,m,c|\hat{\beta}^n) + m\min(s(w^*,r,m,c|\hat{\beta}^n),c) + r),$$

where we suppress the dependence of w^* on r, m, and c to economize on notation.

Now, take r and m as given and ask what c is optimal. We will denote the c that maximizes this objective function given r and m as $c^*(r,m)$. Replacing for s, the objective function becomes
(108)

$$V(r,m,c) = \begin{cases} \ln(w^* - \frac{\hat{\beta}^n w^* - mc - r}{1+\hat{\beta}^n}) + \ln(\frac{\hat{\beta}^n w^* - mc - r}{1+\hat{\beta}^n} + mc + r), & \text{if } c < \underline{c}(\hat{\beta}^n) \\ \ln(w^* - c) + \ln(c(1+m) + r), & \text{if } \underline{c}(\hat{\beta}^n) \le c \le \overline{c}(\hat{\beta}^n) \\ \ln(w^* - \frac{\hat{\beta}^n w^* - r/(1+m)}{1+\hat{\beta}^n}) + \ln((1+m)\frac{\hat{\beta}^n w^* - r/(1+m)}{1+\hat{\beta}^n} + r), & \text{if } \overline{c}(\hat{\beta}^n) < c, \end{cases}$$

53

where we have used the fact that $w^*\hat{\beta}^n > r/(1+m)$, since we have assumed the optimal contracts are matching contracts, to eliminate the $\max$ from the savings function in the third piece of the objective function. The objective function can be rewritten as,

(109)
$$V(r,m,c) = \begin{cases} 2\ln(w^* + mc + r) + \ln(\frac{\hat{\beta}^n}{(1+\hat{\beta}^n)^2}), & \text{if } c < \underline{c}(\hat{\beta}^n) \\ \ln(w^* - c) + \ln(c(1+m) + r), & \text{if } \underline{c}(\hat{\beta}^n) \leq c \leq \overline{c}(\hat{\beta}^n) \\ 2\ln(w^* + r/(1+m)) + \ln(1+m) + \ln(\frac{\hat{\beta}^n}{(1+\hat{\beta}^n)^2}), & \text{if } \overline{c}(\hat{\beta}^n) < c \,. \end{cases}$$

The partial derivative with respect to c is,

(110)
$$V_c(r,m,c) = \begin{cases} 2\frac{\frac{\partial w^*}{\partial c} + m}{w^* + mc + r} & , \text{if } c < \underline{c}(\hat{\beta}^n) \\ \frac{\frac{\partial w^*}{\partial c} - 1}{w^* - c} + \frac{1+m}{c(1+m)+r} & , \text{if } \underline{c}(\hat{\beta}^n) < c < \overline{c}(\hat{\beta}^n) \\ 2\frac{\frac{\partial w^*}{\partial c}}{w^* + r/(1+m)}, & \text{if } \overline{c}(\hat{\beta}^n) < c \end{cases}$$

By the implicit function theorem,

(111)
$$\frac{\partial w^*}{\partial c} = -\frac{m\frac{\partial \bar{s}}{\partial c}}{1 + m\frac{\partial \bar{s}}{\partial w}}.$$

Substituting in for the derivatives of $\bar{s}$, this expression becomes

(112)
$$\frac{\partial w^*}{\partial c} = \begin{cases} -m, & \text{if } c < \overline{c}(\beta^n) \\ \frac{-m\kappa^r}{1 + m[\kappa^n \frac{\beta^n}{1+\beta^n}\mathbf{1}(\beta^n w^* > r/(1+m))]}, & \text{if } \overline{c}(\beta^n) < c < \overline{c}(1) \\ 0, & \text{if } \overline{c}(1) < c, \end{cases}$$

where the first line is the case where c is so low that both types save to it, the last line is the case where neither type saves to the cap, and the middle line is the case where only rationals save to the cap. Notice that this expression turns crucially on $\overline{c}(\beta^n)$, the maximum cap to which a naive worker will save. Let $K \equiv \frac{\kappa^r}{1 + m[\kappa^n \frac{\beta^n}{1+\beta^n}\mathbf{1}(\beta^n w^* > r/(1+m))]}$, where $\frac{\kappa^r}{1+m} < K \leq \kappa^r$, as this term often arises. Note that there is no assumption that $\beta^n w^* > r/(1+m)$ and the naive will actually receive matching contributions in any optimal contract.

Consider first whether $c > \overline{c}(\hat{\beta}^n)$ could be optimal. Replacing for $\frac{\partial w^*}{\partial c}$ from (112), the partial derivative of the objective function in this range is,

(113)
$$V_c(r,m,c) = \begin{cases} -\frac{2mK}{w^* + \frac{r}{1+m}} & , \text{if } \overline{c}(\hat{\beta}^n) < c < \overline{c}(1) \\ 0 & , \text{if } \overline{c}(1) < c. \end{cases}$$

This implies that $c = \overline{c}(\hat{\beta}^n)$ is weakly preferred to all higher caps, and the preference is strict for $\hat{\beta}^n < 1$.

Now consider whether $c < \underline{c}(\hat{\beta}^n)$ could be optimal. The derivative in this range depends on the relation between $\underline{c}(\hat{\beta}^n)$ and $\overline{c}(\beta^n)$:

(114)
$$V_c(r,m,c) = \begin{cases} 0 & , \text{if } c < \min\{\underline{c}(\hat{\beta}^n), \overline{c}(\beta^n)\} \\ \frac{2m}{w^* + mc + r}(1 - K) & , \text{if } \overline{c}(\beta^n) < c < \underline{c}(\hat{\beta}^n) \end{cases}$$

54

Thus, again, a cap at $c = \underline{c}(\hat{\beta}^n)$ is weakly preferred to any lower cap and the preference is strict if $\overline{c}(\beta^n) < \underline{c}(\hat{\beta}^n)$.

It is left to consider caps between $\underline{c}(\hat{\beta}^n)$ and $\overline{c}(\hat{\beta}^n)$, where the naive will anticipate saving just to the cap. Replacing for $\frac{\partial w^*}{\partial c}$ in the middle range of equation (110),

(115) $\qquad V_c(r, m, c) = \begin{cases} (1+m)[\frac{1}{c(1+m)+r} - \frac{1}{w^*-c}] & \text{, if } \underline{c}(\hat{\beta}^n) < c < \overline{c}(\beta^n) \\ \frac{(1+m)}{c(1+m)+r} - \frac{1+mK}{w^*-c} & \text{, if } \max\{\overline{c}(\beta^n), \underline{c}(\hat{\beta}^n)\} < c < \overline{c}(\hat{\beta}^n). \end{cases}$

Consider the ends of this $[\underline{c}(\hat{\beta}^n), \overline{c}(\hat{\beta}^n)]$ range of the cap, where the naive anticipates saving exactly to the cap. At that lowest cap, $\underline{c}(\hat{\beta}^n) = \frac{w^*\hat{\beta}^n - r}{1+\hat{\beta}^n+m}$, the return to increasing the cap is given by the right-hand derivative,

(116)

$V_{c+}(r, m, \underline{c}(\hat{\beta}^n)) = \begin{cases} \frac{(1+m)(1-\hat{\beta}^n)[w^*(1+m)+r]}{(1+\hat{\beta}^n+m)(\underline{c}(\hat{\beta}^n)(1+m)+r)(w^*-\underline{c}(\hat{\beta}^n))} & \text{, if } \underline{c}(\hat{\beta}^n) < \overline{c}(\beta^n) \\ \frac{(1+m)(1-\hat{\beta}^n)[w^*(1+m)+r]}{(1+\hat{\beta}^n+m)(\underline{c}(\hat{\beta}^n)(1+m)+r)(w^*-\underline{c}(\hat{\beta}^n))} + \frac{m}{w^*-\underline{c}(\hat{\beta}^n)}(1-K) & \text{, if } \overline{c}(\beta^n) < \underline{c}(\hat{\beta}^n). \end{cases}$

First, note that if $\overline{c}(\beta^n) < \underline{c}(\hat{\beta}^n)$, or if $\hat{\beta}^n < 1$, then this derivative is strictly positive and $c^*(r, m) > \underline{c}(\hat{\beta}^n)$. Otherwise, this derivative is zero.

Now suppose that $\overline{c}(\beta^n) \geq \underline{c}(\hat{\beta}^n)$ and consider the left-hand and right-hand derivatives at $\overline{c}(\beta^n)$. From the left,

(117)
$$V_{c-}(r, m, \overline{c}(\beta^n)) = \frac{-m-1}{w^* - \frac{\beta^n w^* - r/(1+m)}{1+\beta^n}} + \frac{1+m}{(1+m)\frac{\beta^n w^* - r/(1+m)}{1+\beta^n} + r}$$
$$= [(w^* + r/(1+m))\frac{\beta^n}{1+\beta^n}]^{-1}[1 - \beta^n(1+m)],$$

while from the right,

(118)
$$V_{c+}(r, m, \overline{c}(\beta^n)) = \frac{-mK-1}{w^* - \frac{\beta^n w^* - r/(1+m)}{1+\beta^n}} + \frac{1+m}{(1+m)\frac{\beta^n w^* - r/(1+m)}{1+\beta^n} + r}$$
$$= [(w^* + r/(1+m))\frac{\beta^n}{1+\beta^n}]^{-1}[1 - \beta^n(1+mK)].$$

The signs of these derivatives are ambiguous. There are two key cases to consider. First, if $m \geq \frac{\frac{1}{\beta^n}-1}{K}$ ("big m") then $V_{c-}(r, m, \overline{c}(\beta^n)) < 0$ and $V_{c+}(r, m, \overline{c}(\beta^n)) \leq 0$, which implies that $c^*(r, m) < \overline{c}(\beta^n)$ so that both naive and rational workers will actually save enough to receive the full match in the contract. Consider the highest utility achievable in this big m case.

If $\hat{\beta}^n < 1$, any $c^*(r, m) > 0$ must satisfy the first-order condition from equating the first line of equation (115) to zero, which implies that $c^*(r, m) = \max\{\frac{w^*-r}{2+m}, 0\}$. If $r > w^*$, then this is not a matching contract and therefore utility cannot exceed $2\ln(\gamma/2)$. Otherwise, $c^*(r, m) = \frac{w^*-r}{2+m}$, and by zero profits, $w^* = \frac{\gamma}{2}(\frac{2+m}{1+m}) - r$. $r < w^*$ requires $r < \frac{\gamma}{4}(\frac{2+m}{1+m})$. All such contracts deliver anticipated consumption of $c_1 = c_2 = \gamma/2$ and utility of $2\ln(\gamma/2)$.

If $\hat{\beta}^n = 1$, then the derivatives of the objective function above imply that $c^*(r, m)$ includes any $c < \underline{c}(1) = \frac{w^*-r}{2+m}$. If $r > w^*$, then this is not a matching contract and therefore utility cannot exceed $2\ln(\gamma/2)$. Otherwise, $c^*(r, m)(1+m) + r < \gamma/2$ but the naive worker anticipates savings greater than $c^*(r, m)$ in order to equate consumption in the two periods, resulting in utility of $2\ln(\gamma/2)$.

55

Consider for the rest of the proof the second key case, $m < \frac{\frac{1}{\beta^n}-1}{K}$ ("small m"), which implies that $V_{c+}(r,m,\bar{c}(\beta^n)) > 0$. Observe that for any (r,m) such that $c^*(r,m) \leq \bar{c}(\beta^n)$, the best utility that can be achieved is $2\ln(\gamma/2)$. We will show below that there exist (r,m) such that $c^*(r,m) > \bar{c}(\beta^n)$ and that delivers utility strictly greater than $2\ln(\gamma/2)$.

For caps in the interval $(\max\{\bar{c}(\beta^n), \underline{c}(\hat{\beta}^n)\}, \bar{c}(\hat{\beta}^n))$, the derivative of the objective function with respect to c, given by (115), is at first positive and decreasing in c. The left-hand derivative at the top of that interval is,

(119)
$$V_{c-}(r,m,\bar{c}(\hat{\beta}^n)) = \frac{-1-mK}{w^* - \frac{\hat{\beta}^n w^* - r/(1+m)}{1+\hat{\beta}^n}} + \frac{1+m}{(1+m)\frac{\hat{\beta}^n w^* - r/(1+m)}{1+\hat{\beta}^n} + r}$$

$$= [(w^* + r/(1+m))\frac{\hat{\beta}^n}{1+\hat{\beta}^n}]^{-1}[1 - \hat{\beta}^n(1+mK)].$$

This derivative is negative if $m > \frac{\frac{1}{\hat{\beta}^n}-1}{K} \equiv \bar{m}$. In particular, it is negative in any matching contract with $\hat{\beta}^n = 1$, which together with (113) implies that $c^*(r,m) \leq \bar{c}(\hat{\beta}^n)$ for all $\hat{\beta}^n$. If this derivative is negative, then any $c^*(r,m) > 0$ satisfies the first-order condition from equating the second line of equation (115) to zero, which implies that $c^*(r,m) = \max(\frac{w^* - \frac{r}{1+m}(1+mK)}{2+mK}, 0)$. If $c = 0$, then the contract is not a matching contract. Since we have assumed the solution is a matching contract, that implies that $c^*(r,m) = \frac{w^* - \frac{r}{1+m}(1+mK)}{2+mK}$.

Finally, if $m \leq \bar{m}$, so that this derivative is positive, then $c^*(r,m) = \bar{c}(\hat{\beta}^n)$.

Substituting $c^*(r,m)$ into the objective function, the problem becomes,

(120)
$$\max_{r,m} V(r,m) = \ln(w^*(r,m) - c^*(r,m))$$
$$+ \ln(c^*(r,m)(1+m) + r),$$

where $w^*(r,m) \equiv w^*(r,m,c^*(r,m))$. Denote the set of r that maximizes this objective function given m by the set-valued function $r^*(m)$.

Consider the highest utility achievable in each of these small m cases. First, if $m > \bar{m}$ so that $c^*(r,m)$ is in the interior, then the optimization problem becomes,

(121)
$$\max_{r,m} V(r,m) = 2\ln(w^*(r,m) + \frac{r}{1+m}) + \ln(\frac{(1+m)(1+mK)}{(2+mK)^2}).$$

The third line of (103) gives us that $s(w^*, r, m, c^*|\beta^n) = (\frac{\beta^n w^* - \frac{r}{1+m}}{1+\beta^n})\mathbf{1}(w^*\beta^n > \frac{r}{1+m})$. Substituting for c^* and s into the implicit definition of $w^*(r,m)$ and solving for w^* yields an explicit definition,

(122)
$$w^*(r,m) = \frac{\gamma - \frac{r}{1+m}\left(1 + m - m[\kappa^r(\frac{1+mK}{2+mK}) + \kappa^n(\frac{\mathbf{1}(w^*\beta^n > \frac{r}{1+m})}{1+\beta^n})]\right)}{1 + m[\kappa^r(\frac{1}{2+mK}) + \kappa^n(\mathbf{1}(w^*\beta^n > \frac{r}{1+m})\frac{\beta^n}{1+\beta^n})]}$$

$$= \begin{cases} \frac{\gamma}{1+m[\kappa^r(\frac{1}{2+mK})+\kappa^n(\frac{\beta^n}{1+\beta^n})]} - \frac{r}{1+m}, & \text{if } \frac{\gamma\beta^n}{1+m[\kappa^r(\frac{1}{2+mK})+\kappa^n(\frac{\beta^n}{1+\beta^n})]} > \frac{r(1+\beta^n)}{1+m} \\ \frac{\gamma - \frac{m}{1+m}\kappa^n r}{1+\frac{m\kappa^r}{2+m\kappa^r}} - \frac{r}{1+m}, & \text{otherwise.} \end{cases}$$

Replacing for this wage, the objective function becomes,

(123)
$$V(r,m) = \begin{cases} 2\ln\left(\frac{\gamma}{1+m[\kappa^r(\frac{1}{2+mK})+\kappa^n(\frac{\beta^n}{1+\beta^n})]}\right) + \ln\left(\frac{(1+m)(1+mK)}{(2+mK)^2}\right), & \text{if } \frac{\gamma\beta^n}{1+m[\kappa^r(\frac{1}{2+mK})+\kappa^n(\frac{\beta^n}{1+\beta^n})]} > \frac{r(1+\beta^n)}{1+m} \\ 2\ln\left(\frac{\gamma - \frac{m}{1+m}\kappa^n r}{1+\frac{m\kappa^r}{2+m\kappa^r}}\right) + \ln\left(\frac{(1+m)(1+mK)}{(2+mK)^2}\right), & \text{otherwise.} \end{cases}$$

where we can see by inspection that the naive's anticipated payoff is independent of r, unless r get so high that the myope quits savings ($r > w^*\beta^n(1+m)$), at which point the naive's payoff declines in r. Thus, for interior c^*, any r satisfying $r \leq w^*(1+m)\beta^n$ is consistent with maximization. Using (122), we can write this cutoff explicitly as $r \leq \frac{\gamma(1+m)\frac{\beta^n}{1+\beta^n}}{1+m[\kappa^r(\frac{1}{2+mK})+\kappa^n(\frac{\beta^n}{1+\beta^n})]}$.

What is the optimal m in this case? We will maximize the first line in (123) with respect to m, but first it is convenient to simplify:

$$V(r^*,m) = 2\ln(\gamma) + \ln\left[\frac{(1+m)(1+mK)}{[2+mK+m\kappa^r+m\kappa^n[2+mK]\frac{\beta^n}{1+\beta^n}]^2}\right]$$

$$= 2\ln(\gamma) + \ln\left[\frac{(1+m)(1+mK)}{[2+m\kappa^r+m\kappa^n\frac{2\beta^n}{1+\beta^n}+mK(1+m\kappa^n\frac{\beta^n}{1+\beta^n})]^2}\right]$$

(124)
$$= 2\ln(\gamma) + \ln\left[\frac{(1+m)(\frac{1+m(\kappa^n\frac{\beta^n}{1+\beta^n}+\kappa^r)}{1+m\kappa^n\frac{\beta^n}{1+\beta^n}})}{4[1+m(\kappa^r+\kappa^n\frac{\beta^n}{1+\beta^n})]^2}\right]$$

$$= 2\ln(\gamma/2) + \ln\left[\frac{1+m}{[1+m\kappa^n\frac{\beta^n}{1+\beta^n}][1+m(\kappa^r+\kappa^n\frac{\beta^n}{1+\beta^n})]}\right],$$

where the third equality comes from replacing for $K = \frac{\kappa^r}{1+m\kappa^n\frac{\beta^n}{1+\beta^n}}$. If an optimal contract includes an interior cap, then it also includes a match that satisfies,

(125)
$$\frac{1}{1+m} - \frac{\kappa^n\frac{\beta^n}{1+\beta^n}}{1+m\kappa^n\frac{\beta^n}{1+\beta^n}} - \frac{\kappa^r+\kappa^n\frac{\beta^n}{1+\beta^n}}{1+m[\kappa^r+\kappa^n\frac{\beta^n}{1+\beta^n}]} = 0.$$

You can use the quadratic theorem to find the solution, $m = \sqrt{\frac{1+\beta^n\kappa^r}{\beta^n(\beta^n+\kappa^r)}} - 1 > 0$, which is strictly less than $m^{FB} = \frac{1-\beta^n}{\beta^n}$ and therefore also strictly less than $\frac{\frac{1}{\beta^n}-1}{K}$. If this solution exceeds the bottom of the range of m under consideration, $\frac{\frac{1}{\hat\beta^n}-1}{K}$, then it is indeed the optimal m in that range. If not, then the payoff decreases in m throughout this range.

Consider now the case when $m \leq \bar m$, so that $c^*(r,m) = \bar c(\hat\beta^n) = \frac{\hat\beta^n w^* - r/(1+m)}{1+\hat\beta^n}$,

(126)
$$V(r,m) = 2\ln(w^*(r,m) + \frac{r}{1+m}) + \ln(1+m) + \ln\left(\frac{\hat\beta^n}{(1+\hat\beta^n)^2}\right),$$

and,

$$
w^*(r,m) = \frac{\gamma - \frac{r}{1+m}\left(1 + m - m\left[\frac{\kappa^r}{1+\hat{\beta}^n} + \frac{\kappa^n}{1+\beta^n}\mathbf{1}(w^*(r,m)\beta^n > \frac{r}{1+m})\right]\right)}{1 + m\left[\frac{\kappa^r\hat{\beta}^n}{1+\hat{\beta}^n} + \frac{\kappa^n\beta^n}{1+\beta^n}\mathbf{1}(w^*(r,m)\beta^n > \frac{r}{1+m})\right]}
$$

(127)
$$
= \begin{cases} \frac{\gamma}{1+m\left[\frac{\kappa^r\hat{\beta}^n}{1+\hat{\beta}^n} + \frac{\kappa^n\beta^n}{1+\beta^n}\right]} - \frac{r}{1+m}, & \text{if } \frac{\gamma\beta^n}{1+m\left[\frac{\kappa^r\hat{\beta}^n}{1+\hat{\beta}^n} + \frac{\kappa^n\beta^n}{1+\beta^n}\right]} > \frac{r(1+\beta^n)}{1+m} \\ \frac{\gamma - \frac{m}{1+m}\kappa^n r}{1+m\frac{\kappa^r\hat{\beta}^n}{1+\hat{\beta}^n}} - \frac{r}{1+m}, & \text{otherwise.} \end{cases}
$$

Replacing for this wage,

(128)
$$
V(r,m) = \begin{cases} 2\ln\left(\frac{\gamma}{1+m\left[\frac{\kappa^r\hat{\beta}^n}{1+\hat{\beta}^n} + \frac{\kappa^n\beta^n}{1+\beta^n}\right]}\right) + \ln(1+m) + \ln\left(\frac{\hat{\beta}^n}{(1+\hat{\beta}^n)^2}\right), & \text{if } \frac{\gamma\beta^n}{1+m\left[\frac{\kappa^r\hat{\beta}^n}{1+\hat{\beta}^n} + \frac{\kappa^n\beta^n}{1+\beta^n}\right]} > \frac{r(1+\beta^n)}{1+m} \\ 2\ln\left(\frac{\gamma - \frac{m}{1+m}\kappa^n r}{1+m\frac{\kappa^r\hat{\beta}^n}{1+\hat{\beta}^n}}\right) + \ln(1+m) + \ln\left(\frac{\hat{\beta}^n}{(1+\hat{\beta}^n)^2}\right), & \text{otherwise.} \end{cases}
$$

Again, in this case, the payoff is independent of r, up to a point (where $s(w^*(r,m), r, m, c^*(r,m)|\beta^n) = 0$), and then decreases in r. As before, we can solve for this cutoff explicitly, and derive $r \leq \frac{\gamma(1+m)\frac{\beta^n}{1+\beta^n}}{1+m\left[\frac{\kappa^r\hat{\beta}^n}{1+\hat{\beta}^n} + \frac{\kappa^n\beta^n}{1+\beta^n}\right]}$.

What is the optimal m in this range? It is convenient to define $\bar{\sigma} \equiv \left[\kappa^r\frac{\hat{\beta}^n}{1+\hat{\beta}^n} + \kappa^n\frac{\beta^n}{1+\beta^n}\right]$, the average savings *rate* in the contract if $r = 0$. The first-order condition, defined by setting the partial derivative of the first line of (128) with respect to m to zero, can be written as,

(129)
$$
\frac{1}{1+m} - \frac{2\bar{\sigma}}{1+m\bar{\sigma}} = 0,
$$

which reduces to $m = \frac{1-2\bar{\sigma}}{\bar{\sigma}} < m^{FB} < \frac{m^{FB}}{K}$. If this solution exceeds the top of the range of m under consideration, $\frac{\frac{1}{\hat{\beta}^n}-1}{K}$, then it is indeed the optimal m in that range. If not, then the optimal m in that range is equal to $\frac{\frac{1}{\hat{\beta}^n}-1}{K}$.

We have now shown that in each of these two subcases of the small m case, the optimal m is less than m^{FB}. We also showed above that the highest utility achievable in the big m case is $2\ln(\gamma/2)$. We now show that the utility achievable in the small m case is strictly greater than $2\ln(\gamma/2)$. To see this, consider the contract with $c = \bar{c}(\hat{\beta}^n)$, $r = 0$, $m = \frac{1-2\bar{\sigma}}{\bar{\sigma}}$, and $w = w^*(r,m,c)$. Our results above show that this contract yields the naives utility of,

(130)
$$
V(r,m) = 2\ln(\gamma/2) + \ln\left(\frac{\hat{\beta}^n}{(1+\hat{\beta}^n)^2}\right) - \ln(\bar{\sigma}(1-\bar{\sigma})) > 2\ln(\gamma/2),
$$

where the inequality comes from the fact that $x(1-x)$ is increasing when $x < 1/2$ and $\bar{\sigma} < \frac{\hat{\beta}^n}{1+\hat{\beta}^n} < 1/2$. And so the best contract in the small m case must achieve utility strictly higher than $2\ln(\gamma/2)$. This implies that any optimal contract must be in the small m case and must have $m < m^{FB}$, $c > \max(\underline{c}(\hat{\beta}^n), \bar{c}(\beta^n))$, and $c \leq \bar{c}(\hat{\beta}^n)$. This implies that any optimal contract uses a matching cap equal to naives' anticipated savings level such that naives save strictly less than the cap.

It is left to prove that the rationals' payoff under the optimal contract is the same as naives' payoff. Since we know that any optimal contract must offer an m in one of the two small m

subcases, we will consider each subcase in turn, beginning with the interior case, when $c^*(r, m) <$
$\bar{c}(\hat{\beta}^n) \leq \bar{c}(1)$. Since payoffs are independent of r, for all r that are in used in any optimal contract
(see equation (123)), it suffices to consider the case when $r = 0$. In that case, $c^*(0, m) = \frac{w^*}{2+mK} >$
$\frac{w^*}{2+m} = \underline{c}(1)$, so that the rational saves just to the cap. Since naives and rationals anticipate saving
the same amount, they receive the same payoff strictly greater than $2\ln(\gamma/2)$.

For the $c^*(r, m) = \bar{c}(\hat{\beta}^n)$ subcase, note that $\underline{c}(1) > \bar{c}(\hat{\beta}^n)$ whenever $\frac{\hat{\beta}^n w - \frac{r}{1+m}}{1+\hat{\beta}^n} < \frac{w-r}{2+m}$. Here too
it suffices to consider the case where $r = 0$ (see equation 123), where the rational saves just to the
cap in any optimal contract whenever,

$$
\begin{aligned}
\underline{c}(1) < \bar{c}(\hat{\beta}^n) &\iff \frac{w^*}{2+m} < \frac{w^* \hat{\beta}^n}{1+\hat{\beta}^n} \\
&\iff \frac{1}{2+m} < \frac{\hat{\beta}^n}{1+\hat{\beta}^n} \\
&\iff 2+m > \frac{1+\hat{\beta}^n}{\hat{\beta}^n}
\end{aligned}
$$

(131)

Suppose m in an optimal contract is in the interior of this subcase, so that $m = \frac{1}{\bar{\sigma}} - 2$. Then for
this to be true we must have $\frac{1}{\bar{\sigma}} > \frac{1+\hat{\beta}^n}{\hat{\beta}^n}$, which is always true. Suppose instead there is an optimal
contract with m at the boundary between the two subcases, $\frac{\frac{1}{\hat{\beta}^n} - 1}{K}$. Then we have,

$$
\begin{aligned}
\underline{c}(1) < \bar{c}(\hat{\beta}^n) &\iff 2 + \frac{\frac{1}{\hat{\beta}^n} - 1}{K} > \frac{1+\hat{\beta}^n}{\hat{\beta}^n} \\
&\iff 2\hat{\beta}^n + \frac{1-\hat{\beta}^n}{K} > 1 + \hat{\beta}^n \\
&\iff \frac{1-\hat{\beta}^n}{K} > 1 - \hat{\beta}^n,
\end{aligned}
$$

(132)

which is always true. Hence in either type of optimal contract, rationals save to the cap and hence
naives and rationals anticipate savings the same amount and therefore receive the same payoff,
strictly greater that $2ln(\gamma/2)$.

$\square$

Let C^*_{rn} represent the set of contracts preferred by the naives from the set of all nonnegative-
profit contacts when pooled with the rationals. Consider an equilibrium consisting of a subset of
C^*_{rn}, preferred by the naives and rationals, and the contract $(\gamma/2, \gamma/2, 0, 0)$, preferred by the so-
phisticates. We will show that such a set of contracts is an equilibrium. If not, there must be some
alternative contract (w, r, m, c) that one or more types strictly prefer to their equilibrium contracts
that would make nonnegative profits. Clearly, no alternative contract can be strictly preferred
by sophisticates or rationals, on their own, and make nonnegative profits since no nonnegative-
profit separating contract can deliver a total compensation above γ, yielding utility no better than
$2\ln(\gamma/2)$. Furthermore, no nonnegative-profit pooling contract can be strictly preferred by the
sophisticates, since they can never receive total consumption exceeding γ, even in a pooling con-
tract, since they will always save weakly less than average. Finally, there can be no alternative
nonnegative-profit pooling contract strictly preferred by both naives and rationals by the definition

of C_{rn}^*. Thus, the only type of alternative contract that remains to be considered is one that is strictly preferred by only the naives. But no such entrant can exist. We proved in lemma 9 that the naives and rationals get the same payoff in the proposed equilibrium. In any contract, rationals' payoff under the contract is weakly greater than naives' payoff. (To see this, note that what naives' self 0 believes she will save under the contract is a choice available to the rationals' self 1. Since rationals' self 0 and self 1 share the same objective function as naives' self 0, this means that rationals must always do at least as well as naives.) Thus any contract preferred by the naives to the equilibrium contract will also be preferred by the rationals. This proves that any set of contracts consisting of a subset of C_{rn}^* and the contract $(\gamma/2, \gamma/2, 0, 0)$ is an equilibrium.

Finally, any equilibrium contract set must include a subset of C_{rn}^* preferred by naives and rationals and one or more contracts that deliver $c_1 = c_2 = \gamma/2$ to sophisticates and no other contracts.

To see this, first, note that by the same argument as in the proof of Lemma 3, in any equilibrium the sophisticates receive a payoff of $2\ln(\gamma/2)$. The only way for them to receive that in equilibrium is through a contract that delivers $c_1 = c_2 = \gamma/2$. Note as well that contracts in C_{rn}^* do not deliver consumption of $c_1 = c_2 = \gamma/2$ to sophisticates since they offer a matching rate less than m^{FB}.

Second, in any equilibrium, naives must receive a payoff at least as good as they receive in the contracts in the set C_{rn}^*, which we showed above is strictly greater than $2\ln(\gamma/2)$. Suppose not. Then a contract in the set C_{rn}^* could enter, make naives strictly better off, and make nonnegative profits given that naives and (potentially) rationals would strictly prefer it, which is a contradiction with the definition of equilibrium.

Third, in any equilibrium, rationals must pool with naives. Suppose not. By the same argument as in the proof of Lemma 4, the best the rationals can do in a separating contract is $2\ln(\gamma/2)$. However, we have already shown that naives must receive an equilibrium payoff at least as high as they receive in contracts in the set C_{rn}^*, which is greater than $2\ln(\gamma/2)$. In any contract, rationals' payoff under the contract is weakly greater than naives' payoff. This implies that rationals would strictly prefer the contract preferred by naives, which is a contradiction.

Fourth, rationals and naives must pool in contracts in the set C_{rn}^*. Suppose not. Then, since we have already shown that rationals must pool with naives, that implies that rationals and naives pool in some set of contracts not in C_{rn}^*. But then a contract in the set C_{rn}^* could enter, make naives strictly better off (by definition of C_{rn}^*), and make nonnegative profits given that naives and (potentially) rationals would choose it, which is a contradiction with the definition of equilibrium. □

Proof of Proposition 5.

Lemma 10. *Any equilibrium matching contract must employ the default that minimizes average matched savings under the contract, given the other contractual terms. That default is either 0 or $\underline{d}$.*

Proof. Suppose the proposition is not true and there is some competitive equilibrium in which some matching contract (w, r, m, c, d) includes a default that does not minimize average matched savings. First, suppose that contract is preferred by rationals and naives. Consider the alternative contract (w', r, m, c, d') that uses the default d' that minimizes matched savings and a slightly higher wage $w' > w$. That alternative contract would be strictly preferred by both rationals and naives. Moreover, there exists a $w' > w$ such that that alternative contract would still make nonnegative profits since lower matching payments are made under the alternative contract. This implies a failure of the free entry condition of equilibrium.

Second, suppose the equilibrium contract (w, r, m, c, d) is preferred by only one type, either naives or rationals. This implies that the other type strictly prefers some other contract in the equilibrium. This implies that there exists an alternative contract (w', r, m, c, d') that is strictly preferred by the type that prefers (w, r, m, c, d) from among the equilibrium contracts and that the other type does not prefer to its preferred equilibrium contracts, and that makes nonnegative profits. This implies a failure of the free entry condition of equilibrium.

To see that the default that minimizes average matched savings in a matching contract is either 0 or $\underline{d}$, first note that, relative to $d = 0$, any default in $(0, \underline{d})$ or greater than $\bar{d}$ results in the same savings outcomes for advice takers and active savers but strictly higher savings for procrastinators. Similarly, relative to $d = \underline{d}$, any default in $(\underline{d}, \bar{d}]$ results in the same savings outcomes for active savers but strictly higher savings for procrastinators and advice takers.

□

Lemma 11. *The highest utility rationals can achieve from among the set of all contracts that make nonnegative profits when taken up by only the rationals exceeds $2 \ln(\gamma/2)$ and approaches $2 \ln(\gamma/2)$ as κ^a approaches 1.*

Proof. Assume for now that any contract in the constraint set that delivers the highest utility to rationals—which we will refer to as an "optimal contract" for short—is a contract in which the worker anticipates receiving matching contributions (a "matching contract"), so that $m \min(s(w, r, m, c), c) > 0$. We will show later that the payoff from the best matching contacts exceeds $2 \ln(\gamma/2)$, which is the highest payoff possible in a non-matching contract.

The constrained optimization problem is:

(133)
$$\max_{w,r,m,c,d} V(w, r, m, c, d) = \ln(w - s(w, r, m, c)) +$$
$$\ln(s(w, r, m, c) + m \min(s(w, r, m, c), c) + r),$$

subject to,

(134)
$$w = \gamma - \bar{s}(w, r, m, c, d)m - r,$$

where,

(135)
$$s(w, r, m, c) = \begin{cases} \frac{w - mc - r}{2} & \text{, if } c < \underline{c} \\ c & \text{, if } \underline{c} \leq c \leq \bar{c} \\ \max\{\frac{w - r/(1+m)}{2}, 0\} & \text{, if } \bar{c} < c, \end{cases}$$

with $\bar{c} = \frac{w - r/(1+m)}{2}$ and $\underline{c} = \frac{w - r}{2 + m}$, and where,

$$\bar{s}(w, r, m, c, d) = \begin{cases} \left[\kappa^a + \kappa^t\right] \min(c, s(w, r, m, c)) + \kappa^p \min(c, dw) & \text{if} \quad d < \underline{d} \\ \kappa^a \min(c, s(w, r, m, c)) + \left[\kappa^p + \kappa^t\right] \min(c, dw) & \text{if} \quad d \in [\underline{d}, \bar{d}] \\ \min(c, s(w, r, m, c)) & \text{if} \quad d > \bar{d}. \end{cases}$$

Non-satiation implies that we can restrict attention to zero profit contracts; hence the nonnegative profit constraint (134) holds with equality.

The savings level of active savers given in (135) comes from the function in equation (103) evaluated at $\beta = 1$,

An optimal contract must use the default that minimizes average matched savings under the contract, given the other contractual terms, following the argument in the proof of lemma 10. This implies that we only need to consider $d \in \{0, \underline{d}\}$.

61

Implicitly define the zero-profit wage as a function of the other contract parameters using (134):

(136) $$w^*(r, m, c, d) = \gamma - \bar{s}(w^*(r, m, c, d), r, m, c, d)m - r.$$

The optimization problem can now be rewritten as,

(137)
$$\max_{r,m,c,d} V(r, m, c, d) = \ln(w^*(r, m, c, d) - s(w^*(r, m, c, d), r, m, c)) +$$
$$\ln(s(w^*(r, m, c, d), r, m, c) + m\min(s(w^*(r, m, c, d), r, m, c), c) + r),$$

Substituting in for $s(w^*(r, m, c, d), r, m, c)$, the objective function becomes,

(138)
$$V(r, m, c, d) = \begin{cases} 2\ln(\frac{w^* + mc + r}{2}), & \text{if } c < \underline{c} \\ \ln(w^* - c) + \ln((1+m)c + r), & \text{if } \underline{c} < c < \bar{c} \\ 2\ln(\frac{w^* + r/(1+m)}{2}) + \ln(1+m), & \text{if } \bar{c} < c, \end{cases}$$

where we suppress the arguments of w^* to economize on notation and where we use our assumption that any optimal contract is a matching contract to eliminate the $\max$ from the savings function in the third piece of the objective function.

Taking the partial derivative with respect to c,

(139)
$$V_c(r, m, c, d) = \begin{cases} \frac{2(m + \frac{\partial w^*}{\partial c})}{2 + mc + r}, & \text{if } c < \underline{c} \\ \frac{\frac{\partial w^*}{\partial c} - 1}{w^* - c} + \frac{1 + m}{(1+m)c + r}, & \text{if } \underline{c} < c < \bar{c} \\ 2\frac{\frac{\partial w^*}{\partial c}}{w^* + r/(1+m)}, & \text{if } \bar{c} < c. \end{cases}$$

By the implicit function theorem,

(140)
$$\frac{\partial w^*}{\partial c} = -\frac{m\frac{\partial \bar{s}}{\partial c}}{1 + m\frac{\partial \bar{s}}{\partial w}}.$$

Substituting for savings,

(141)
$$\frac{\partial w^*}{\partial c} = \begin{cases} -\frac{m\kappa^a}{1 + m(1 - \kappa^a)\underline{d}}, & \text{if } c < \bar{c} \text{ and } d = \underline{d} \text{ with } \underline{d}w^* < c \\ -m(\kappa^a + \kappa^t), & \text{if } c < \bar{c} \text{ and } d = 0 \\ 0, & \text{if } \bar{c} < c \text{ and either } d = 0 \text{ or } d = \underline{d} \text{ with } \underline{d}w^* < c. \end{cases}$$

In the interest of brevity, in the piecewise function above we do not consider the $d = \underline{d}$ case with $\underline{d}w^* > c$. The reason is that if $\underline{d}w^* > c$, then $d = \underline{d}$ is not the matched-savings-minimizing default since workers who stay at the default would then receive at least as much matching contributions as active savers, and therefore matched savings would be lower if $d = 0$. For the rest of the proof, whenever we consider the case with $d = \underline{d}$, we also assume $\underline{d}w^* < c$ without mentioning it every time.

By inspection of equation (139), together with the calculation in (141), we can rule out $c < \underline{c}$ and notice indifference over all caps above $\bar{c}$. Thus, consider caps in the $[\underline{c}, \bar{c}]$ range. Let $K(d) = \frac{\kappa^a}{1 + m(1 - \kappa^a)\underline{d}}$ if $d = \underline{d}$ and $K(d) = \kappa^a + \kappa^t$ if $d = 0$. Then the derivative of the value function with respect to c in this range of c is

(142)
$$V_c(r, m, c, d) = -\frac{1 + mK(d)}{w^* - c} + \frac{1 + m}{(1+m)c + r}.$$

In particular, at $\underline{c} = \frac{w^* - r}{2+m}$,

(143) $$V_{c+}(r, m, \underline{c}, d) = (\frac{1+m}{2+m})[\frac{w^* m(1 - K(d)) + \frac{r}{1+m} + K(d)r}{(w^* - \underline{c})((1+m)\underline{c} + r)}] > 0.$$

At $\overline{c} = \frac{w^* - \frac{r}{1+m}}{2}$,

(144) $$V_{c-}(r, m, \overline{c}, d) = -(\frac{m(1+m)K(d)}{2})[\frac{w^* - \frac{r}{1+m}}{(w^* - \overline{c})((1+m)\overline{c} + r)}] < 0,$$

where we know that $w^* > r/(1 + m)$ by our initial assumption that this is a matching contract since otherwise $s(w^*, r, m, c) = 0$.

Denote the optimal c, given r, m, and d, as $c^*(r, m, d)$. Taken together, these facts imply that $\underline{c} < c^*(r, m, d) < \overline{c}$, and $c^*(r, m, d)$ satisfies the first-order condition,

(145) $$\frac{1 + mK(d)}{w^* - c^*} = \frac{1 + m}{(1 + m)c^* + r}$$

or $c^* = \frac{w^* - (1+mK(d))\frac{r}{1+m}}{2 + mK(d)}$. Substituting this expression for c^* for c in the value function in (138), we can rewrite the optimization problem as,

(146) $$\max_{r,m,d} V(r, m, d) = 2\ln(w^*(r, m, d) + \frac{r}{1+m}) + \ln(\frac{1 + mK(d)}{(2 + mK(d))^2}) + \ln(1 + m)$$

where,

(147) $$w^*(r, m, d) \equiv w^*(r, m, c^*(r, m, d), d) = \begin{cases} \frac{(2+mK(0))(\gamma - r) + mK(0)(1+mK(0))\frac{r}{1+m}}{2(1+mK(0))}, & \text{if } d = 0 \\ \frac{\gamma - r + r\kappa^a \frac{m}{1+m} \frac{1+mK(\underline{d})}{2+mK(\underline{d})}}{1 + m[\frac{\kappa^a}{2+mK(\underline{d})} + (1-\kappa^a)\underline{d}]}, & \text{if } d = \underline{d} \end{cases}$$

Taking the partial derivative of the value function in the problem (146) with respect to r,

(148) $$V_r(r, m, d) = \frac{\frac{\partial w^*}{\partial r} + \frac{1}{1+m}}{w^*(r, m, d) + \frac{r}{1+m}}$$
$$= \begin{cases} \frac{mK(0)(1+mK(0)) + 2(1+mK(0)) - (2+mK(0))(1+m)}{2(1+mK(0))(1+m)[w^*(r,m,0) + \frac{r}{1+m}]}, & \text{if } d = 0 \\ \frac{\kappa^a m \frac{1+mK(\underline{d})}{2+mK(\underline{d})} + 1 + m[\frac{\kappa^a}{2+mK(\underline{d})} + (1-\kappa^a)\underline{d}] - (1+m)}{[1 + m[\frac{\kappa^a}{2+mK(\underline{d})} + (1-\kappa^a)\underline{d}]](1+m)[w^*(r,m,\underline{d}) + \frac{r}{1+m}]}, & \text{if } d = \underline{d} \end{cases}$$
$$= \begin{cases} \frac{2m(K(0)-1) + m^2 K(0)(K(0)-1)}{2(1+mK(0))(1+m)[w^*(r,m,0) + \frac{r}{1+m}]} < 0, & \text{if } d = 0 \\ \frac{m(\kappa^a + (1-\kappa^a)\underline{d}) - m}{[1 + m[\frac{\kappa^a}{2+mK(\underline{d})} + (1-\kappa^a)\underline{d}]](1+m)[w^*(r,m,\underline{d}) + \frac{r}{1+m}]} < 0, & \text{if } d = \underline{d}. \end{cases}$$

Thus, for any m or d, the rational's optimal contract includes $r^*(m, d) = 0$. In a step similar to the replacement for c^*, above, we can now calculate the zero-profit wage.

(149) $$w^*(m, d) \equiv w^*(0, m, d) = \begin{cases} \frac{\gamma}{2}\frac{2 + mK(0)}{1 + mK(0)}, & \text{if } d = 0 \\ \frac{\gamma}{2}\frac{2 + mK(\underline{d})}{1 + m[\kappa^a + (1-\kappa^a)\underline{d}]}, & \text{if } d = \underline{d} \end{cases}$$

and further rewrite the optimization problem

$$\max_{m,d} V(m,d) = 2\ln(w^*(m,d)) + \ln(\frac{1+mK(d)}{(2+mK(d))^2}) + \ln(1+m)$$

(150)
$$= \begin{cases} 2\ln(\frac{\gamma}{2}) + \ln(\frac{1+m}{1+mK(0)}), & \text{if } d = 0 \\ 2\ln(\frac{\gamma}{2}) + \ln(\frac{1+m}{[1+m[\kappa^a+(1-\kappa^a)\underline{d}]][1+m(1-\kappa^a)\underline{d}]}), & \text{if } d = \underline{d} \end{cases}$$

Finally, consider the optimal match. Simplifying and taking the derivative,

(151)
$$V_m(m,d) = \begin{cases} -\frac{(\kappa^a+\kappa^t)}{1+m[\kappa^a+\kappa^t]} + \frac{1}{1+m}, & \text{if } d = 0 \\ -[\frac{(1-\kappa^a)\underline{d}}{1+m(1-\kappa^a)\underline{d}} + \frac{\kappa^a+(1-\kappa^a)\underline{d}}{1+m[\kappa^a+(1-\kappa^a)\underline{d}]}] + \frac{1}{1+m}, & \text{if } d = \underline{d}. \end{cases}$$

Inspection reveals that the payoff in the $d = 0$ contract is strictly increasing in m for all m. Thus, in that case, the payoff can be made arbitrarily close to the limit of $V(m,0)$ as m approaches infinity, which by inspection of the first line of the value function (150) is $2\ln(\gamma/2) - \ln(\kappa^a + \kappa^t)$.

Under an automatic-enrollment strategy with $d = \underline{d}$, the optimal match is positive (assuming $\underline{d} < 1/2$), and satisfies $m^* + 1 = \sqrt{(\frac{1-\sigma_1}{\sigma_1})(\frac{1-\sigma_2}{\sigma_2})}$, where $\sigma_1 = \kappa^a + (1-\kappa^a)\underline{d}$ and $\sigma_2 = (1-\kappa^a)\underline{d}$. Plugging this in to the value function,

(152)
$$V(m^*, \underline{d}) = 2\ln(\gamma/2) - 2\ln[\sqrt{(1-\sigma_1)\sigma_2} + \sqrt{\sigma_1(1-\sigma_2)}]$$
$$= 2\ln(\gamma/2) - 2\ln[(1-\kappa^a)\sqrt{(1-\underline{d})\underline{d}} + \sqrt{\kappa^a + (1-\kappa^a)^2\underline{d}(1-\underline{d})}],$$

which decreases in κ^a, reaching $2\ln(\gamma/2)$ as $\kappa^a \to 1$.

Both these payoffs exceed $2\ln(\gamma/2)$, the best a rational can do in a non-matching contract, so the rational's preferred zero-profit contract is a matching contract of the sort outlined. But these payoffs approach $2\ln(\gamma/2)$ from above at κ^a approaches 1. $\qquad\square$

Lemma 12. *Consider the set of contracts that make nonnegative profits when taken up by both naives and rationals. There exists a $\underline{\kappa} < 1$ such that if $\kappa^a > \underline{\kappa}$ then naives' preferred contracts in that set are matching contracts such that: (1) the matching cap is set at the naives' anticipated savings level; (2) naives who make an active choice in fact save strictly less than the matching cap; (3) rationals who make an active choice save at the matching cap; (4) the default contribution rate is the one that minimizes average worker savings in the contract, given the other terms of the contract, and the default contribution amount, dw, is strictly below the contract's matching cap; and (5) it achieves a payoff for both rationals and naives of at least $2\ln(\gamma/2) + \ln(\frac{\hat{\beta}^n}{(1+\hat{\beta}^n)^2}) - \ln(\bar{\sigma}(1-\bar{\sigma})) > 2\ln(\gamma/2)$, where $\bar{\sigma} = \kappa^n\frac{\beta}{1+\beta} + \kappa^r\frac{\hat{\beta}}{1+\hat{\beta}}$.*

Proof. Note that the naives' preferred contracts in this set must make zero profits since otherwise the contract's wage could be increased to make naives strictly better off while still making non-negative profits. Hence we can restrict attention to contracts that would make zero profits when chosen by only naives and rationals.

Assume for now that any contract in the constraint set that delivers the highest utility to naives—which we will refer to as an "optimal contract" for short—is a contract in which the worker antici-pates receiving matching contributions (a "matching contract"), so that $m\min(s(w,r,m,c,d|\hat{\beta}^n),c) >$

0. We will show that the payoff from the best matching contracts exceeds $2\ln(\gamma/2)$. Since the highest payoff possible in a non-matching contract is $2\ln(\gamma/2)$ (following the argument in the proof of lemma 6), this implies that any optimal contract must be a matching contract.

The constrained optimization problem is:

$$
(153) \quad \max_{w,r,m,c,d} V(w,r,m,c,d) = \ln(w - s(w,r,m,c|\hat{\beta}^n))
$$
$$
+ \ln(s(w,r,m,c|\hat{\beta}^n) + m\min(s(w,r,m,c|\hat{\beta}^n),c) + r),
$$

subject to,

$$
(154) \qquad\qquad w = \gamma - \bar{s}(w,r,m,c,d)m - r,
$$

where,

$$
(155) \qquad s(w,r,m,c|\beta) = \begin{cases} \frac{\beta w - mc - r}{1+\beta} & \text{, if } c < \underline{c}(\beta) \\ c & \text{, if } \underline{c}(\beta) \le c \le \bar{c}(\beta) \\ \max\{\frac{\beta w - r/(1+m)}{1+\beta}, 0\} & \text{, if } \bar{c}(\beta) < c. \end{cases}
$$

with $\bar{c}(\beta) = \frac{w\beta - r/(1+m)}{1+\beta}$ and $\underline{c}(\beta) = \frac{w\beta - r}{1+\beta+m}$, and where,

$$
\bar{s}(w,r,m,c,d) = \begin{cases} \left[\kappa^a + \kappa^t\right]\left[\kappa^r\min(c,s(1)) + \kappa^n\min(c,s(\beta^n))\right] + \kappa^p\min(c,dw) & \text{if } d < \underline{d} \\ \kappa^a\left[\kappa^r\min(c,s(1)) + \kappa^n\min(c,s(\beta^n))\right] + \left[\kappa^p + \kappa^t\right]\min(c,dw) & \text{if } d \in [\underline{d}, \bar{d}] \\ \left[\kappa^r\min(c,s(1)) + \kappa^n\min(c,s(\beta^n))\right] & \text{if } d > \bar{d}, \end{cases}
$$

where we have suppressed the dependence of s on the contract terms to economize on notation.

The proof proceeds in parallel to the proof of lemma 9 from equation (106) to (111), which we will not repeat here. The first difference in the proof introduced by defaults is in the expression for $\frac{\partial w^*}{\partial c}$, which now must take into account the effect of defaults on $\bar{s}$. An optimal contract must use the default that minimizes average matched savings under the contract, given the other contractual terms, following the argument in the proof of lemma 10. This implies that we only need to consider $d \in \{0, \underline{d}\}$. We can thus express $\frac{\partial w^*}{\partial c}$ as,

$$
(156)
$$

$$
\frac{\partial w^*}{\partial c} = \begin{cases} -\frac{m\kappa^a}{1+m(1-\kappa^a)\underline{d}}, & \text{if } c < \bar{c}(\beta^n) \text{ and } d = \underline{d} \text{ with } \underline{d}w^* < c \\ -m(\kappa^a + \kappa^t), & \text{if } c < \bar{c}(\beta^n) \text{ and } d = 0 \\ \frac{-m\kappa^r\kappa^a}{1+m[\kappa^n\kappa^a\frac{\beta^n}{1+\beta^n}\mathbf{1}(\beta^n w^* > r/(1+m))+(1-\kappa^a)\underline{d}]}, & \text{if } d = \underline{d} \text{ with } \underline{d}w^* < c \text{ and } \bar{c}(\beta^n) < c < \bar{c}(1) \\ \frac{-m\kappa^r(\kappa^a+\kappa^t)}{1+m[\kappa^n(\kappa^a+\kappa^t)\frac{\beta^n}{1+\beta^n}\mathbf{1}(\beta^n w^* > r/(1+m))]}, & \text{if } d = 0 \text{ and } \bar{c}(\beta^n) < c < \bar{c}(1) \\ 0, & \text{if } \bar{c}(1) < c \text{ and either } d = 0 \text{ or } d = \underline{d} \text{ with } \underline{d}w^* < c, \end{cases}
$$

where as in (141), we do not need to consider the $d = \underline{d}$ case with $\underline{d}w^* > c$.

Denote the optimal c, given r, m, and d, as $c^*(r,m,d)$. Consider first whether $c > \bar{c}(\hat{\beta}^n)$ could be optimal. By inspection of V_c (given in (110) above), together with the calculation in (156), $c = \bar{c}(\hat{\beta}^n)$ is weakly preferred to all higher caps, for either default choice, and the preference is strict for $\hat{\beta}^n < 1$.

Now consider whether $c < \underline{c}(\hat{\beta}^n)$ could be optimal. The calculation in (156) shows that $|\frac{\partial w^*}{\partial c}| < m$, for either default choice. Given this fact, inspection of V_c in (110) allows us to rule out any cap below $\underline{c}(\hat{\beta}^n)$, for either default choice, as the naive's objective function strictly increases in the cap.

It is left to consider caps between $\underline{c}(\hat{\beta}^n)$ and $\bar{c}(\hat{\beta}^n)$, where the naive will anticipate saving just to the cap. In this range,

$$(157) \qquad V_c(r, m, c, d) = \frac{(1+m)}{c(1+m)+r} - \frac{1+mK(d,c)}{w^* - c}$$

where $K(d,c) \equiv \frac{-1}{m}\frac{\partial w^*}{\partial c}$ is strictly between zero and 1 and depends on both the default and whether the cap exceeds $\bar{c}(\beta^n)$.

Consider first the lowest cap in this range, $\underline{c}(\hat{\beta}^n)$. The right-hand derivative there is given by,

$$
\begin{aligned}
V_{c+}(r, m, \underline{c}(\hat{\beta}^n), d) &= \frac{(1+m)}{\frac{w^*\hat{\beta}^n - r}{1+\hat{\beta}^n+m}(1+m)+r} - \frac{1+mK(d,\underline{c}(\hat{\beta}^n))}{w^* - \frac{w^*\hat{\beta}^n - r}{1+\hat{\beta}^n+m}} \\
(158) \qquad &= \frac{\frac{1}{\hat{\beta}^n}(1+m) - (1+mK(d,\underline{c}(\hat{\beta}^n)))}{\frac{w^*(1+m)+r}{1+\hat{\beta}^n+m}} > 0,
\end{aligned}
$$

and therefore $c^*(r, m, d) > \underline{c}(\hat{\beta}^n)$.

Now suppose $\bar{c}(\beta^n) > \underline{c}(\hat{\beta}^n)$. For $\underline{c}(\hat{\beta}^n) < c < \bar{c}(\beta^n)$, V_c is strictly decreasing in c. It is also strictly decreasing in c when $\bar{c}(\beta^n) < c < \bar{c}(\hat{\beta}^n)$. The derivatives from the right and left at $\bar{c}(\beta^n)$ are:

$$(159) \qquad V_{c+}(r, m, \bar{c}(\beta^n), d) = \frac{1 - (1 + mK(d, \bar{c}(\beta^n)_+))\beta^n}{\frac{\beta^n}{1+\beta^n}[w^* + \frac{r}{1+m}]}$$

$$(160) \qquad V_{c-}(r, m, \bar{c}(\beta^n), d) = \frac{1 - (1 + mK(d, \bar{c}(\beta^n)_-))\beta^n}{\frac{\beta^n}{1+\beta^n}[w^* + \frac{r}{1+m}]},$$

where $K(d, c_+)$ and $K(d, c_-)$ are the values of $K(d, c)$, on the right and left side of c, respectively.

Note that, for given $d \in \{0, \underline{d}\}$, $K(d, c_+) < K(d, c_-)$, so the derivative from the right is greater than the derivative from the left. Thus, if the derivative from the left is positive, then so is the derivative from the right. In general, the signs of these derivatives are ambiguous, turning crucially on m.

First, if $m \geq \frac{\frac{1}{\beta^n}-1}{K(d,\bar{c}(\beta^n)_+)}$ ("big m"), then $V_{c-}(r, m, \bar{c}(\beta^n), d) < 0$ and $V_{c+}(r, m, \bar{c}(\beta^n), d) \leq 0$, which implies that $c^*(r, m, d) < \bar{c}(\beta^n)$ so that both naive and rational workers will actually save enough to receive the full match in the contract. It is easy to bound the highest payoff for naives possible in this subset of the contract space. To do so, note that the set of nonnegative-profit contracts that satisfy $c^*(r, m, d) < \bar{c}(\beta^n)$ considered here is a strict subset of the constraint set considered in lemma 11, and moreover rationals' payoff from any given contract is always weakly greater than naives' payoff. Therefore, the highest payoff possible in this big m case must be weakly less than the highest payoff possible to rationals in the constraint set considered in lemma 11, which approaches $2ln(\gamma/2)$ as κ^a approaches 1.

Consider for the rest of the proof the second key case, $m < \frac{\frac{1}{\beta^n}-1}{K(d,\bar{c}(\beta^n)_+)}$ ("small m"), which implies that $V_{c+}(r, m, \bar{c}(\beta^n), d) > 0$ so that $c^*(r, m, d) > \bar{c}(\beta^n)$, and if $c^*(r, m, d) < \bar{c}(\hat{\beta}^n)$ it must

satisfy $V_c(r, m, c, d) = 0$, where V_c is defined as in (157) and,

$$(161) \qquad K(d, c) = K(d) = \begin{cases} \frac{\kappa^r \kappa^a}{1 + m[\kappa^n \kappa^a \frac{\beta^n}{1+\beta^n} \mathbf{1}(\beta^n w^* > r/(1+m)) + (1-\kappa^a)\underline{d}]}, & \text{if } d = \underline{d} \text{ with } \underline{d}w^* < c \\ \frac{\kappa^r (\kappa^a + \kappa^t)}{1 + m[\kappa^n (\kappa^a + \kappa^t) \frac{\beta^n}{1+\beta^n} \mathbf{1}(\beta^n w^* > r/(1+m))]}, & \text{if } d = 0, \end{cases}$$

where we can drop the dependence on c since in the range of caps under consideration it no longer depends on c.

At $\bar{c}(\hat{\beta}^n) = \frac{w^*\hat{\beta}^n - r\frac{r}{1+m}}{1+\hat{\beta}^n}$,

$$(162) \qquad V_{c-}(r, m, \bar{c}(\hat{\beta}^n), d) = \frac{1 - \hat{\beta}^n(1 + mK(d))}{\frac{\hat{\beta}^n}{1+\hat{\beta}^n}(w + \frac{r}{1+m})}.$$

This derivative is negative if $m > \frac{\frac{1}{\hat{\beta}^n} - 1}{K(d)}$. In particular, it is negative in any matching contract with $\hat{\beta}^n = 1$, which together with (156) and (156) implies that $c^*(r, m, d) \leq \bar{c}(\hat{\beta}^n)$ for all $\hat{\beta}^n$. If this derivative is negative, then any $c^*(r, m, d) > 0$ satisfies the first-order condition,

$$(163) \qquad \frac{1 + mK(d)}{w^* - c^*} = \frac{1 + m}{(1+m)c^* + r}$$

or $c^*(r, m, d) = \frac{w^* - (1+mK(d))\frac{r}{1+m}}{2+mK(d)}$. If $c = 0$, then the contract is not a matching contract, and since we have assumed the solution is a matching contract, that implies that this first-order condition is satisfied in any optimal contract. Finally, if $m \leq \frac{\frac{1}{\hat{\beta}^n} - 1}{K(d)}$, then this derivative is positive and $c^*(r, m, d) = \bar{c}(\hat{\beta}^n)$.

Substituting $c^*(r, m, d)$ into the objective function, the problem becomes,

$$(164) \qquad \max_{r,m,d} V(r, m, d) = \ln(w^*(r, m, d) - c^*(r, m, d)) \\ + \ln(c^*(r, m, d)(1 + m) + r),$$

where $w^*(r, m, d) \equiv w^*(r, m, c^*(r, m, d), d)$. Denote the r that maximizes this objective function given m and d by the potentially set-valued function $r^*(m, d)$.

Consider the optimal r in each of these small m cases. First, if $m > \frac{\frac{1}{\hat{\beta}^n} - 1}{K(d)}$ so that $c^*(r, m, d)$ is in the interior, then the optimization problem becomes,

$$(165) \qquad \max_{r,m,d} V(r, m, d) = 2\ln(w^*(r, m, d) + \frac{r}{1+m}) + \ln(\frac{(1+m)(1 + mK(d))}{(2 + mK(d))^2}).$$

The third line of (155) gives us that $s(w^*, r, m, c^*, d|\beta^n) = (\frac{\beta^n w^* - \frac{r}{1+m}}{1+\beta^n})\mathbf{1}(w^*\beta^n > \frac{r}{1+m})$. Substituting for c^* and s into the implicit definition of $w^*(r, m, d)$ and solving for w^* yields an explicit

definition,

(166)

$$
w^* =
\begin{cases}
\dfrac{\gamma-\frac{r}{1+m}\left(1+m-m(\kappa^a+\kappa^t)[\kappa^r(\frac{1+mK(0)}{2+mK(0)})+\kappa^n(\frac{\mathbf{1}(w^*\beta^n>\frac{r}{1+m})}{1+\beta^n})]\right)}{1+m(\kappa^a+\kappa^t)[\kappa^r(\frac{1}{2+mK(0)})+\kappa^n(\mathbf{1}(w^*\beta^n>\frac{r}{1+m})\frac{\beta^n}{1+\beta^n})]}, & \text{if } d=0 \\[2em]
\dfrac{\gamma-\frac{r}{1+m}\left(1+m-m\kappa^a[\kappa^r(\frac{1+mK(\underline{d})}{2+mK(\underline{d})})+\kappa^n(\frac{\mathbf{1}(w^*\beta^n>\frac{r}{1+m})}{1+\beta^n})]\right)}{1+m\kappa^a[\kappa^r(\frac{1}{2+mK(\underline{d})})+\kappa^n(\mathbf{1}(w^*\beta^n>\frac{r}{1+m})\frac{\beta^n}{1+\beta^n})]+m(1-\kappa^a)\underline{d}}, & \text{if } d=\underline{d}
\end{cases}
$$

$$
=
\begin{cases}
\dfrac{\gamma-\frac{mr}{1+m}\kappa^p}{1+m(\kappa^a+\kappa^t)[\frac{\kappa^r}{2+mK(0)}+\kappa^n\frac{\beta^n}{1+\beta^n}]}-\frac{r}{1+m}, & \text{if } d=0 \text{ and } \frac{\beta^n(\gamma-\kappa^p\frac{rm}{1+m})}{1+m(\kappa^a+\kappa^t)[\frac{\kappa^r}{2+mK(0)}+\kappa^n\frac{\beta^n}{1+\beta^n}]}>\frac{r(1+\beta^n)}{1+m} \\[1.5em]
\dfrac{\gamma-\frac{mr}{1+m}((\kappa^a+\kappa^t)\kappa^n+\kappa^p)}{1+m(\kappa^a+\kappa^t)[\frac{\kappa^r}{2+mK(0)}]}-\frac{r}{1+m}, & \text{if } d=0 \text{ and } \frac{\beta^n(\gamma-\kappa^p\frac{rm}{1+m})}{1+m(\kappa^a+\kappa^t)[\frac{\kappa^r}{2+mK(0)}+\kappa^n\frac{\beta^n}{1+\beta^n}]}\leq\frac{r(1+\beta^n)}{1+m} \\[1.5em]
\dfrac{\gamma-\frac{mr}{1+m}(1-\kappa^a)(1-\underline{d})}{1+m\kappa^a[\frac{\kappa^r}{2+mK(\underline{d})}+\kappa^n\frac{\beta^n}{1+\beta^n}]+m(1-\kappa^a)\underline{d}}-\frac{r}{1+m}, & \text{if } d=\underline{d} \text{ and } \frac{\beta^n(\gamma-\frac{mr}{1+m}(1-\kappa^a)(1-\underline{d}))}{1+m\kappa^a[\frac{\kappa^r}{2+mK(\underline{d})}+\kappa^n\frac{\beta^n}{1+\beta^n}]+m(1-\kappa^a)\underline{d}}>\frac{r(1+\beta^n)}{1+m} \\[1.5em]
\dfrac{\gamma-\frac{mr}{1+m}(\kappa^a\kappa^n+(1-\kappa^a)(1-\underline{d}))}{1+m\kappa^a[\frac{\kappa^r}{2+mK(\underline{d})}]+m(1-\kappa^a)\underline{d}}-\frac{r}{1+m}, & \text{if } d=\underline{d} \text{ and } \frac{\beta^n(\gamma-\frac{mr}{1+m}(1-\kappa^a)(1-\underline{d}))}{1+m\kappa^a[\frac{\kappa^r}{2+mK(\underline{d})}+\kappa^n\frac{\beta^n}{1+\beta^n}]+m(1-\kappa^a)\underline{d}}\leq\frac{r(1+\beta^n)}{1+m}
\end{cases}
$$

In all four cases $w^* + \frac{r}{1+m}$ is strictly decreasing in r, and therefore the objective function in (165) is strictly decreasing in r, which implies that $r^*(m,d)=0$.

Consider now the case where $m \leq \frac{\frac{1}{\hat{\beta}^n}-1}{K(d)}$, so that $c^*(r,m,d)=\bar{c}(\hat{\beta}^n)=\frac{\hat{\beta}^n w^*-r/(1+m)}{1+\hat{\beta}^n}$. Following the approach taken for the $c^*(r,m,d)<\bar{c}(\hat{\beta}^n)$ case from equations (165) to (166), it is easy to show that the objective function is also decreasing in r in this case, so that $r^*(m,d)=0$ in this range of m as well. We can thus write $w^*(r^*,m,d)$ as,

(167) $\qquad w^*(m,d) \equiv w^*(r^*,m,d) =
\begin{cases}
\dfrac{\gamma}{1+m(\kappa^a+\kappa^t)(\kappa^r\frac{\hat{\beta}^n}{1+\hat{\beta}^n}+\kappa^n\frac{\beta^n}{1+\beta^n})}, & \text{if } d=0 \\[1.5em]
\dfrac{\gamma}{1+m(\kappa^a(\kappa^r\frac{\hat{\beta}^n}{1+\hat{\beta}^n}+\kappa^n\frac{\beta^n}{1+\beta^n})+(1-\kappa^a)\underline{d})}, & \text{if } d=\underline{d},
\end{cases}$

and,

$$V(m,d) \equiv V(r^*,m,d) = \ln(w^*(m,d)(1-\frac{\hat{\beta}^n}{1+\hat{\beta}^n})) + \ln((1+m)w^*(m,d)\frac{\hat{\beta}^n}{1+\hat{\beta}^n})$$

(168)
$$= 2\ln(w^*) + \ln((1-\frac{\hat{\beta}^n}{1+\hat{\beta}^n})\frac{\hat{\beta}^n}{1+\hat{\beta}^n}) + \ln(1+m)$$

$$= 2\ln(\gamma) + \ln((1-\frac{\hat{\beta}^n}{1+\hat{\beta}^n})\frac{\hat{\beta}^n}{1+\hat{\beta}^n}) + \ln(1+m) - 2\ln(1+m\sigma(d)),$$

where,

(169) $\qquad \sigma(d) =
\begin{cases}
(\kappa^a+\kappa^t)(\kappa^r\frac{\hat{\beta}^n}{1+\hat{\beta}^n}+\kappa^n\frac{\beta^n}{1+\beta^n}), & \text{if } d=0 \\[1em]
\kappa^a(\kappa^r\frac{\hat{\beta}^n}{1+\hat{\beta}^n}+\kappa^n\frac{\beta^n}{1+\beta^n})+(1-\kappa^a)\underline{d}, & \text{if } d=\underline{d}
\end{cases}$

is the average savings rate in a contract with default d. It is a simple maximization to show that the optimal matching rate is $\bar{m}^*(d)=\frac{1-2\sigma(d)}{\sigma(d)}$.

We showed above that as κ^a goes to 1, the highest utility achievable in the big m case goes to $2\ln(\gamma/2)$. We now show that as κ^a goes to 1, the utility achievable in the small m case asymptotes to a level strictly greater than $2\ln(\gamma/2)$. To see this, consider the contract with $c=\bar{c}(\hat{\beta}^n)$, $r=0$,

$m = \frac{1 - 2\sigma(0)}{\sigma(0)}$, $d = 0$ and $w = w^*(r, m, c, d)$. Our results above show that this contract yields the naives utility of,

$$(170) \qquad V(r, m) = 2\ln(\gamma/2) + \ln\left(\frac{\hat{\beta}^n}{(1 + \hat{\beta}^n)^2}\right) - \ln(\sigma(0)(1 - \sigma(0))).$$

As κ^a goes to 1, this payoff asymptotes from above to the analogous payoff from the case without defaults considered in lemma 9, given in (130), which is strictly greater than $2\ln(\gamma/2)$. This implies that there exists a $\underline{\kappa}$ such that for all $\kappa^a > \underline{\kappa}$, any optimal contract must be in the small m case and must have $c > \max(\underline{c}(\hat{\beta}^n), \bar{c}(\beta^n))$, and $c \le \bar{c}(\hat{\beta}^n)$ so that the matching cap equals naives' anticipated savings level and naives in fact save strictly less than the cap.

It is left to prove that for sufficiently large $\kappa^a < 1$, the rationals' payoff under the optimal contract is the same as naives' payoff. We have shown that, for sufficiently large $\kappa^a < 1$, any optimal contract must offer an m in one of the two small m subcases. Following the same argument as at the end of the proof of lemma 9, the c in any optimal contract must be strictly greater than $\underline{c}(1)$. Hence for sufficiently large $\kappa^a < 1$, in any optimal contract, rationals save to the cap and hence naives and rationals anticipate saving the same amount and therefore receive the same payoff strictly greater than $2\ln(\gamma/2)$.

Finally, following the argument in the proof of lemma 10, the default of any optimal contract must minimize matched savings under the contract, given the other terms of the contract. In particular, any optimal contract must have $dw < c$, since otherwise $d = 0$ would reduce matched savings. Note that for sufficiently large $\kappa^a < 1$, in any optimal contract all savings are matched and thus the default must also minimize average savings under the contract.

$\square$

Let C_{rn}^* represent the set of contracts preferred by the naives from the set of all nonnegative-profit contacts when pooled with the rationals. Consider an equilibrium consisting of a subset of C_{rn}^* in which each equilibrium contract is preferred by both naives and rationals. We will show that if κ^a is sufficiently close to 1, then such a set of contracts is an equilibrium. If not, there must be some alternative contract (w, r, m, c, d) that is strictly preferred by one or more types and would make nonnegative profits if chosen by the types that strictly prefer it.

First, note that by lemmas 11 and 12, for κ^a sufficiently close to 1, the highest payoff rationals could achieve in an alternative nonnegative-profit contract preferred by only rationals is below their payoff from a contract in C_{rn}^*. So there cannot be an alternative nonnegative-profit contract strictly preferred by only rations. Second, there can be no nonnegative-profit pooling contract that is strictly preferred by both naives and rationals by the definition of C_{rn}^*. Thus, the only type of alternative contract that remains to be considered is one that is strictly preferred by only the naives. But no such entrant can exist. We proved in lemma 12 that the naives and rationals get the same payoff in the proposed equilibrium. In any contract, rationals' payoff under the contract is weakly greater than naives' payoff. Thus any contract preferred by the naives to the equilibrium contract will also be preferred by the rationals.

Finally, any equilibrium must consist of a subset of C_{rn}^*.

To see this, first, note that, in any equilibrium, naives must receive a payoff at least as high as they receive in the contracts in the set C_{rn}^*, which we showed above is strictly greater than $2\ln(\gamma/2)$. Suppose not. Then a contract in the set C_{rn}^* could enter, make naives strictly better off, and make nonnegative profits given that naives and (potentially) rationals would choose it, which is a contradiction with the definition of equilibrium.

Second, if $\kappa^a < 1$ is sufficiently high, then in any equilibrium, rationals must pool with naives. Suppose not. Then by lemma 11, we can choose κ^a close enough to 1 to make rationals' highest payoff in a nonnegative-profit separating contract arbitrarily close to $2\ln(\gamma/2)$. However, we have already shown that naives must receive an equilibrium payoff at least as high as they receive in contracts in the set $C_{rn}^{\prime*}$, which is greater than $2\ln(\gamma/2)$. In any contract, rationals' payoff under the contract is weakly greater than naives' payoff. This implies that rationals would strictly prefer the contract preferred by naives, which is a contradiction.

Third, rationals and naives must pool in contracts in the set C_{rn}^*. Suppose not. Then a contract in the set C_{rn}^* could enter, make naives strictly better off, and make nonnegative profits given that naives and (potentially) rationals would choose it, which is a contradiction with the definition of equilibrium.

Finally, note that by lemma 12, contracts in C_{rn}^* have all of the characteristics asserted in the statement of the proposition.

$\square$

Proof of Proposition 6. It will prove useful to denote the fees that induce each type to invest outside the plan by $f_j^O \equiv 1 - r_f$ for $j \in \{R, N\}$. Consider first workers' savings choices in period 1. Facing investment management fees of $\hat{f}$, workers choose s to solve,

$$(171) \qquad \max_s \ln(w - s) + \ln((1 - \hat{f})s).$$

The first-order condition gives the solution,

$$(172) \qquad s^* = \frac{1}{2}w.$$

Rationals' utility from a contract (w, f) is thus given by $\ln(w - \frac{1}{2}w) + \ln((1 - f^1)\frac{1}{2}w)$ if $f^1 \leq f_R^O$, so that they invest inside the plan, and $\ln(w - \frac{1}{2}w) + \ln((1 - f_R^O)\frac{1}{2}w)$ if $f^1 > f_R^O$, so that they invest using their outside option. Naives' utility from a contract (w, f) is simply w. Profits from a contract (w, f) are a function only of w and of f^1 and $\bar{f}$, given workers' investment behavior. Note then that any set of investment options f is equivalent to any other set of investment options f' if $\bar{f} = \bar{f}'$ and $f^1 = f'^1$, because they then generate the same utility and profits. In the analysis that follows we therefore focus on those two key dimensions of f.

Consider the possibility of a pooling equilibrium (i.e., an equilibrium in which all equilibrium contracts are preferred by both types). For any particular set of investment options in a pooling contract, it will be useful to characterize the wage such that the contract would earn zero profits. For investment options f, if both types invest inside the plan, the zero-profit wage is given by $w = \gamma + \frac{1}{2}[\kappa_R f^1 + (1 - \kappa_R)\bar{f}]w$. Solving for w, we have that the zero-profit wage as a function of f is given by,

$$(173) \qquad w(f) = \frac{\gamma}{1 - \frac{1}{2}[\kappa_R f^1 + (1 - \kappa_R)\bar{f}]},$$

for $\bar{f} \leq f_N^O$ and $f^1 \leq f_R^O$. This wage is increasing in $\bar{f}$ up to $\bar{f} = f_N^O$, above which the zero-profit wage falls since naives invest in their outside option.

This fact implies that for any contract (w, f) in a pooling equilibrium, we must have $\bar{f} = f_N^O$. To see why, suppose there were an equilibrium contract (w, f) with $\bar{f} < f_N^O$. Then an alternative contract (w', f') exists with $\bar{f}' > \bar{f}$, $f'^1 = f^1$, and $w' > w$ that makes both types strictly better off and still makes nonnegative profits. Similarly, if $\bar{f} > f_N^O$, then there exists an alternative contract

(w', f') with $\bar{f}' = f_N^O$ and $w' > w$ that makes both types strictly better off and makes nonnegative profits. The existence of such alternative contracts contradicts the definition of equilibrium, so we must have $\bar{f} = f_N^O$.

Consider a contract (w, f) in such a pooling equilibrium and suppose first that f^1 is such that rationals save using their outside option, which requires that $f^1 > f_R^O$. To satisfy the free entry condition of equilibrium, the pooling contract must earn zero profits, which requires $w = \gamma + (1 - \kappa_R)\frac{1}{2}f_N^O w$. Solving for w, we get $w = \frac{\gamma}{1 - \frac{1}{2}(1-\kappa_R)f_N^O}$.

For this to be an equilibrium, there cannot be an alternative contract that would make some type strictly better off and make nonnegative profits. Consider in particular an alternative pooling contract (w', f') in which rationals save within the plan. This requires that $f'^1 \leq f_R^O$. To make this entering contract as attractive to both types as possible, assume it offers $\bar{f}' = f_N^O$ and would make zero profits if both rationals and naives chose it, which requires that $w' = \frac{\gamma}{1 - \frac{1}{2}[\kappa_R f'^1 + (1-\kappa_R)f_N^O]}$.

Note that for $f'^1 > 0$ this wage is higher than the wage of our proposed pooling equilibrium contract, so naives would prefer this entering contract. Rationals would also prefer this entering contract since it both pays a higher wage and charges a lower f^1 than the supposed equilibrium contract. This implies that there cannot be a pooling equilibrium in which rationals save using their outside option.

Suppose instead then that $f^1 \leq f_R^O$ so that rationals save inside the plan. For this to be an equilibrium, there cannot be an alternative pooling contract that would make both naives and rationals strictly better off and make nonnegative profits. This implies that (w, f) must earn zero profits, which requires that $w = \frac{\gamma}{1 - \frac{1}{2}[\kappa_R f^1 + (1-\kappa_R)f_N^O]}$. Naives' equilibrium utility (given simply by w) is therefore monotonically increasing in f^1. Rationals' equilibrium utility in contrast is given by,

(174)
$$\ln(\frac{1}{2}\frac{\gamma}{1 - \frac{1}{2}[\kappa_R f^1 + (1-\kappa_R)f_N^O]}) + \ln(\frac{1}{2}\frac{\gamma(1 - f^1)}{1 - \frac{1}{2}[\kappa_R f^1 + (1-\kappa_R)f_N^O]})$$
$$= 2\ln(\frac{\gamma}{2}) - 2\ln(1 - \frac{1}{2}[\kappa_R f^1 + (1-\kappa_R)(f_N^O)]) + \ln(1 - f^1).$$

Taking the partial derivative of this with respect to f^1 and setting it to zero, we get,

(175)
$$\frac{\kappa^R}{1 - \frac{1}{2}[\kappa_R f^1 + (1-\kappa_R)f_N^O]} - \frac{1}{1 - f^1} = 0,$$

or,

(176)
$$f^1 = 2 - \frac{2}{\kappa_R} + \frac{1 - \kappa_R}{\kappa_R}f_N^O.$$

We must have $f^1 \in [0, f_R^O]$ for it to be feasible and induce rationals to invest inside the plan. For parameter values for which the value of f^1 from the first-order condition (176) is negative, rationals' optimal pooling contract has $f^1 = 0$. For parameter values for which the value of f^1 from the first-order condition (176) is greater than f_R^O, rationals' optimal pooling contract has $f^1 = f_R^O$. Interior solutions are also possible. Denote the f^1 from rationals' optimal zero-profit pooling contract as f^{1*}. Any pooling equilibrium must offer an $f^1 \in [f^{1*}, f_R^O]$, since otherwise an alternative pooling contract exists that would make both types better off and make nonnegative profits. Such an alternative pooling contract does not exist so long as $f^1 \in [f^{1*}, f_R^O]$.

All that remains to consider to determine whether this is an equilibrium is whether there is an alternative contract that would make only one of the types strictly better off and would make

nonnegative profits. Consider first whether there is an alternative contract (w', f') that would make naives and not rationals strictly better off. To make it as unattractive to rationals as possible and to raise as much fee revenue as possible (to fund a high wage), let $f'^1 = \bar{f}' = f_N^O$. To make naives strictly better off, we must have $w' > w$. To not make rationals strictly better off, we must have,

$$\text{(177)} \qquad \ln(\frac{1}{2}w') + \ln((1 - f_R^O)\frac{1}{2}w') \le \ln(\frac{1}{2}w) + \ln((1 - f^1)\frac{1}{2}w),$$

or,

$$\text{(178)} \qquad 2\ln(\frac{1}{2}w') - 2\ln(\frac{1}{2}w) \le \ln(1 - f^1) - \ln(1 - f_R^O).$$

First, suppose $f^1 < f_R^O$. The RHS of (178) is then strictly positive. By continuity of the log function, there thus exists a $w' > w$ such that (178) is satisfied. For this alternative contract to foil our proposed equilibrium, it must also make nonnegative profits. The equilibrium contract makes zero-profits:

$$\text{(179)} \qquad \gamma - w + \frac{1}{2}w(\kappa_R f^1 + (1 - \kappa_R)f_N^O) = 0.$$

This implies that,

$$\text{(180)} \qquad \gamma - w + \frac{1}{2}w f_N^O > 0.$$

Since the LHS of this inequality is a continuous function of w, there must exist a $w' > w$ such that $\gamma - w' + \frac{1}{2}w' f_N^O > 0$, so that the alternative contract makes nonnegative profits. This implies that we cannot have an equilibrium pooling contract with $f^1 < f_R^O$. The only remaining possibility is $f^1 = f_R^O$. This is a pooling equilibrium in which the rationals pay the maximal fee possible and still invest inside the plan. At such a fee level, there is not an alternative contract that makes naives strictly better off and rationals not strictly better off. To make naives strictly better off, such an alternative contract must have $w' > w$. But then it must also make rationals strictly better off since they cannot be made to pay a higher fee than in the equilibrium pooling contract.

Finally, consider whether there is an alternative contract (w', f') that could enter and make rationals and not naives strictly better off and make nonnegative profits. Rationals' most preferred contract among those that earn nonnegative profits when taken only by rationals is the solution to,

$$\text{(181)} \qquad \max_{f^1 \in [0, f_R^O]} \ln(\frac{1}{2}\frac{\gamma}{1 - \frac{f^1}{2}}) + \ln((1 - f_L)\frac{1}{2}\frac{\gamma}{1 - \frac{f^1}{2}}),$$

where we have substituted in the zero-profit wage. It is easy to see that the solution is $f^{1'} = 0$, which implies a zero-profit wage of $w' = \gamma$, so the contract is not preferred by naives to the equilibrium pooling contract. Rationals' utility in such a contract is $2\ln(\frac{\gamma}{2})$.

This contract will not be strictly preferred by rationals to the equilibrium pooling contract if and only if,

$$\text{(182)} \qquad 2\ln(\gamma/2) \le \ln(\frac{1}{2}\frac{\gamma}{1 - \frac{1}{2}[\kappa_R f_R^O + (1 - \kappa_R)f_N^O]}) + \ln(\frac{1}{2}\frac{\gamma(1 - f_R^O)}{1 - \frac{1}{2}[\kappa_R f_R^O + (1 - \kappa_R)f_N^O]}).$$

(182) is thus a necessary and sufficient condition for existence of a pooling equilibrium. We can reduce (182) to,

$$\text{(183)} \qquad [1 - \frac{1}{2}[\kappa_R f_R^O + (1 - \kappa_R)f_N^O]]^2 \le 1 - f_R^O.$$

This condition holds as long as f_R^O is not too large and f_N^O is sufficiently greater than f_R^O. The condition given in the statement of Proposition 6 assumes that this inequality holds strictly. If it only held weakly, then there would also exist semi-separating equilibria, analyzed below. We have also shown that all contracts in a pooling equilibrium must offer $\bar{f} = f_N^O$, $f^1 = f_R^O$ and $w = \frac{\gamma}{1 - \frac{1}{2}[\kappa_R f_R^O + (1 - \kappa_R) f_N^O]}$.

Consider now the possibility of an equilibrium in which there is no contract preferred by both types, i.e., a separating equilibrium. Suppose a separating equilibrium exists in which rationals prefer a contract (w_R, f_R) and naives prefer a contract (w_N, f_N). This requires each type to strictly prefer their contract to the other type's contract. It also requires each contract to make nonnegative profits. This implies the following four inequalities:

(184)
$$w_N > w_R,$$

(185)
$$\ln(\frac{1}{2} w_R) + \ln((1 - f_R^1)\frac{1}{2} w_R) > \ln(\frac{1}{2} w_N) + \ln((1 - f_N^1)\frac{1}{2} w_N),$$

(186)
$$\gamma - w_N + \bar{f}_N \frac{1}{2} w_N \geq 0,$$

and,

(187)
$$\gamma - w_R + f_R^1 \frac{1}{2} w_R \geq 0.$$

For this to be an equilibrium, it must also satisfy free entry. As we showed above, rationals' most preferred contract among those that earn nonnegative profits when taken only by rationals has $f^1 = 0$ and $w = \gamma$. Rationals' utility in such a contract is $2 \ln(\frac{\gamma}{2})$. In any separating equilibrium, free entry implies that rationals must receive this contract, since otherwise this contract could enter, make rationals strictly better off, and make nonnegative profits (even if naives' also prefer it, since they pay weakly higher fees). So we must have $w_R = \gamma$ and $f_R^1 = 0$.

Similarly, naives' most preferred contract among those that earn nonnegative profits when taken only by naives has $\bar{f} = f_N^O$ and $w = \frac{\gamma}{1 - \frac{1}{2} f_N^O}$. In any separating equilibrium, naives must receive that contract. To see this, suppose there were a separating equilibrium in which naives received some $w_N < \frac{\gamma}{1 - \frac{1}{2} f_N^O}$. By definition of a separating equilibrium, we know that,

(188)
$$\ln(\frac{1}{2} w_R) + \ln((1 - f_R^1)\frac{1}{2} w_R) > \ln(\frac{1}{2} w_N) + \ln((1 - f_N^1)\frac{1}{2} w_N).$$

But this implies that an alternative contract with $w' > w_N$ and $\bar{f}' = f_N^O$ could enter and make naives and not rationals strictly better off while making nonnegative profits (since it would not attract rationals). But that is a contradiction with this being an equilibrium. So we must have $w_N \geq \frac{\gamma}{1 - \frac{1}{2} f_N^O}$. But the only such contracts that satisfy that and make nonnegative profits are those with $w_N = \frac{\gamma}{1 - \frac{1}{2} f_N^O}$ and $\bar{f}_N = f_N^O$.

This pair of contracts thus satisfies free entry—there does not exist an alternative contract that would make one or both types strictly better off and make nonnegative profits. The only one of the four inequalities (184) - (187) that is not obviously satisfied by this pair of contracts is rationals' IC constraint, (185). Setting $f_N^1 \geq f_R^O$ makes this easiest to satisfy without affecting any other

constraints or equilibrium payoffs. Rationals' IC constraint now becomes,

$$(189) \qquad 2\ln(\frac{\gamma}{2}) > \ln(\frac{1}{2}\frac{\gamma}{1 - \frac{1}{2}f_N^O}) + \ln((1 - f_R^O)\frac{1}{2}\frac{\gamma}{1 - \frac{1}{2}f_N^O}).$$

Note that our condition for a pooling equilibrium to exist, (182), which is implied by the condition assumed in Proposition 6, implies that (189) does not hold and hence that there cannot be a separating equilibrium.

Finally, consider the possibility of a semi-separating equilibrium in which there is one set of pooling contracts preferred by both types and another set of contracts preferred by only one of the types. Denote an equilibrium pooling contract by (w_P, f_P), a naive-only contract by (w_N, f_N), and a rational-only contract by (w_R, f_R).

We must have $w_N = w_P$ to maintain indifference by naives. Furthermore, to satisfy free entry, we must have $\bar{f}_N = f_N^O$ and $w_N = \frac{\gamma}{1 - \frac{1}{2}f_N^O}$. But that implies that (w_P, f_P) would make negative profits, since the pooling contract would collect strictly less fees but have to pay the same wage. This is a contradiction. This implies that there cannot be any naive-only contracts in a semi-separating equilibrium.

For the rationals-only contracts, we must have $f_R^1 = 0$ and $w_R = \gamma$ in order to satisfy free entry. For rationals to be indifferent between the pooling contracts and the rationals-only contracts, we must have,

$$(190) \qquad \ln(\frac{1}{2}w_P) + \ln((1 - f_P^1)\frac{1}{2}w_P) = 2\ln(\frac{\gamma}{2}).$$

This set of contracts, however, would fail the free entry condition of equilibrium. Consider in particular the alternative contract offered in the pooling equilibrium characterized above, with $\bar{f} = f_N^O$, $f^1 = f_R^O$ and $w = \frac{\gamma}{1 - \frac{1}{2}[\kappa_R f_R^O + (1 - \kappa_R)f_N^O]}$, which makes zero profits if it attracts rationals and naives in proportion to their population shares. The condition given the statement of Proposition 6 implies that the payoff to rationals given on the RHS of (182) is strictly greater than $2\ln(\frac{\gamma}{2})$. Moreover, since (w_P, f_P) provides strictly lower utility to rationals than the pooling equilibrium contract, it must offer a strictly lower wage than the pooling equilibrium contract, since the pooling equilibrium contract already charges rationals their maximal fee. This implies that naives would also be made strictly better off by the pooling equilibrium contract. This implies that there cannot be a semi-separating equilibrium. $\qquad \square$