

Social Security[†]

By PETER DIAMOND*

I frequently find economists who express a view of the system that is very far from mine. For example, many young economists and economics students say that they expect to get no benefits at all from Social Security. This expectation does not seem sensible to me. If there is no legislation changing Social Security, trust fund assets and payroll tax revenue (and revenue from the taxation of benefits) are projected to be sufficient to pay all the benefits scheduled under current law until 2042 (Board of Trustees of Social Security and Medicare, 2003). After the trust fund assets are exhausted the payroll tax revenue would continue to be available to pay benefits, with the flow of revenues at that time sufficient to pay roughly three-quarters of the benefits scheduled in current law. The estimate for the end of the 75-year projection period shows enough revenue to pay roughly two-thirds of scheduled benefits. With initial benefits indexed to earnings, average real benefits would be higher than today, although replacement rates would only be roughly 60 percent of current levels for the medium worker. This projection is a far cry from no benefits.

Moreover, I anticipate that Congress will act before the trust fund is exhausted, both lowering

benefits relative to those scheduled under current law and providing additional revenues to finance higher benefits than are payable after 2042. After all, the financial problem of Social Security is not so very large (unlike the larger and more complex set of financial problems of Medicare and Medicaid).¹ An increase in tax revenue of just over 15 percent of currently projected payroll tax revenues would handle the projected cash flow problem for 75 years on a present value basis. On an annual cash flow basis, the share of GDP needed to provide all of the benefits scheduled in current law would increase from 4.4 percent of GDP today to 7.0 percent in 2077.² Like almost everyone else, I do not favor addressing the projected deficit by simply adding more revenues with no other changes. Nor do I picture that solution as having any political prospects. But solving the problem with a mix of benefit reductions and revenue increases does not require large changes, nor does it require a fundamental restructuring of the program.

It is not just in the perception of the projections and the forecast of politics that I find myself in disagreement with opinions that I often hear. More generally, I think the system works better than many economists think. I hope to convince you that the approach inherent in the current U.S. system broadly makes good sense. In particular, I will argue that it makes sense to mandate taxes to finance a reasonable replacement of earnings after retirement; that it makes sense to mandate that retirement benefits be paid as an annuity; that it makes sense to mandate protection for family members, both young children and surviving spouses; that it

[†] Presidential Address delivered at the one hundred fifteenth meeting of the American Economic Association, January 4, 2004, San Diego, CA.

* Department of Economics, E52-344, Massachusetts Institute of Technology, 50 Memorial Drive, Cambridge, MA 02142 (e-mail: pdiamond@MIT.EDU). I am grateful to Robert Hall for suggesting a framework for approaching Social Security for this address and to him and Henry Aaron, Alan Auerbach, Abhijit Banerjee, Nicholas Barr, Olivier Blanchard, Jeffrey Brown, Courtney Coile, Dora Costa, Thomas Davidoff, Esther Duflo, Jonathan Gruber, David Laibson, Jeffrey Liebman, Assar Lindbeck, Alicia Munnell, Peter Orszag, James Poterba, Virginia Reno, Bernard Saffran, Eytan Sheshinski, Robert Solow, and Peter Temin for valuable comments on previous drafts, and to the National Science Foundation for financial support under Grant No. SES0239380. The address draws heavily on the thinking behind and the content of my book with Peter R. Orszag (2004). I have greatly enjoyed and profited from the collaboration on this and earlier work.

¹ I also note that the financial problems of Social Security are smaller than the problems in many other countries, a comparison that seems to surprise some people.

² To put this change in historic context: Spending on public education when the baby boomers were children increased by 2.8 percent of GDP over 25 years; defense spending during the Cold War increased by 2.5 percent of GDP in 2 years; defense spending in the 1980's increased by 1.5 percent of GDP in 6 years; defense spending in the 1990's decreased by 2.6 percent of GDP in 10 years.

makes sense to have a progressive benefit formula; that it makes sense to limit benefits to those who are old enough and stop working or are even older (whether they stop or not); and that having been generous to early cohorts, it makes sense now to continue with a system that is only partially funded.

This is not to say that I agree with all of the details of the current structure by any means. Of course we should change benefit and tax rules so that we restore actuarial balance—so that projected revenues are sufficient to pay for projected benefits over at least 75 years. And other changes would be desirable as well. I am just arguing that the overall design of Social Security makes good sense. In addition to presenting the basis of my support for the broad structure of Social Security, I will identify some rules needing change and I will speculate on why some economists seem to have a different view from mine.

But I will not get into the debate of whether there should be fully funded individual accounts financed from existing payroll tax revenues (carve-out accounts). Nor shall I discuss the political and economic issues associated with the potential role of stocks in Social Security. Those controversial subjects would take up most of the address and I prefer to write about more fundamental issues. Those interested in my view as to why carve-out accounts would not be good policy in the United States today can turn to Chapter 8 in my book with Peter R. Orszag (2004), which also contains a package of changes to restore actuarial balance and strengthen protection of some vulnerable groups. We discuss the potential role of stocks as well.

I will not say much about the advantage of a mandate to save for retirement—there is little call for eliminating such a mandate. After a discussion of a framework for thinking about Social Security (Section I), I will consider annuitization (Section II), treatment of the family (Section III), the interaction among income distribution, insurance, and labor supply (Section IV), the degree of funding (Section V), and adjustments over time to benefits and taxes (Section VI).³ Not discussed but worth keeping

in mind are the supporting antipoverty programs as they affect the elderly [Supplemental Security Income (SSI) and Medicaid], and the provision of medical insurance for the elderly (Medicare). Nor will I discuss the Disability Insurance program, which is a critical part of Social Security.

I. Providing Retirement Income

One-third of the elderly received at least 90 percent of their income from Social Security in 2001, with nearly two-thirds receiving at least half (Social Security Administration, 2003). Yet Social Security was always meant to be a foundation for retirement income and not a level to be relied on exclusively. The average new award for a retired worker in 2002 was just over \$900 per month. For a worker retiring in 2002 at age 62 (the modal retirement age), a worker in the middle of the earnings distribution received a benefit of roughly one-third of (wage-indexed) lifetime average earnings in 2000 dollars. If the worker had a nonworking spouse of the same age, the benefit would be larger—about one-half of the worker's lifetime average earnings.⁴ These are low replacement rates—you would not want to retire on one-third to one-half of what you had earned on average in your lifetime. Benefits would look even lower compared to earnings over the last decade of work for a worker with the typical age-earnings profile. As a foundation for retirement income, Social Security is something substantial to build on. As a level to live on, it is clearly inadequate. Excessive reliance on Social Security, despite its relatively low replacement rates, together with a more general picture of many workers with inadequate wealth at retirement age, seem

Lindbeck and Mats Persson (2003). They refer to the three dimensions as the distinction between defined benefit and defined contribution, funded and unfunded, and actuarial and nonactuarial. All three are matters of degree, not zero-one choices. A primary difference is that my presentation is focused on issues particularly salient in the U.S. context, while I think that theirs was influenced by the Swedish reform.

⁴ Of women receiving benefits in 2002, roughly one-third received benefits solely as beneficiaries. The rest had at least ten years of earnings history. For the latter group, the replacement rate for the couple would be lower than the one reported for the case of a nonearning spouse.

³ My approach has some similarity to the three-dimensional analysis of different pension systems in Assar

the best evidence for evaluating whether workers make adequate preparation for retirement.

I prefer the term “inadequate preparation” to “insufficient savings.” Preparation involves multiple decisions. Indeed, one decision is to save, to have less consumption than after-tax earned income. Another is investing well. A third is getting adequate insurance for earnings risk to have a satisfactory outcome in retirement despite a possibly adverse earnings experience. And a fourth is using insurance to arrange income flows after retirement. There are lots of ways that workers could end up with inadequate consumption after retirement relative to what might sensibly and efficiently be done with earlier earnings.

In addition to having low savings, many workers have problems converting savings in different years into retirement incomes in later years in different states of nature. We know from 401(k) studies that many workers do not diversify sensibly and many do not choose a sensible portfolio for long-term investments.⁵ The tendency of many workers to accept the default allocation set by their employers is suggestive that they do not have a clear view of how to choose a portfolio. Outside employer-organized retirement savings, others pay advisors as much as 1 percent of assets each year to help them select mutual funds (in what are called wrap accounts). Paying 1 percent extra per year reduces the accumulation at the end of a 40-year career by roughly 20 percent.⁶ Mutual funds, even very similar ones, come with quite different annual charges. While the average of charges of mutual funds containing equities is currently 1 ¼ percent of assets per year (including a prorating of front loads), some workers pay much more. A fee of the average size takes away roughly 25 percent from what would be there at retirement without any fee. Thus many workers find it harder to accumulate enough for retirement than they might, than an idealized theory says they should. To be clear, I am not proposing that these market opportunities be banned—although improvement in regulation would be welcome. Rather, I am saying that analysts of Social Security should be realistic

about the actual functioning of the market alternatives.

Investing is only part of the story. We lack market institutions to provide good insurance for the risk in earnings trajectories, thereby affecting the realized pattern of assets at retirement relative to earnings potential. In the Arrow-Debreu framework, workers have deterministic budget constraints from selling their labor supplies conditional on all the states of nature in which they have labor that they choose to sell. That is, they transfer resources across states of nature to those where the purchasing power is needed more. Making the same point in a finance vocabulary, markets do not currently exist for directly hedging the risks in earnings opportunities, and if they did exist I do not think we would see many workers using them.⁷

In addition to problems in converting earnings opportunities into wealth devoted to retirement consumption, the wealth that is privately allocated to retirement consumption does not make adequate use of annuities. This problem would be more severe without the annuities provided by Social Security, since the utility value of the marginal annuity decreases with the extent of existing annuitization.

These shortcomings in providing for retirement income fall on surviving spouses even more heavily than on workers. While 5 percent of elderly married couples have incomes below the poverty line, with another 3 percent near poverty, these figures more than triple when we consider widows. Indeed, widowhood is associated with a roughly 30 percent drop in income relative to needs (Karen Holden and Cathleen Zick, 1998).⁸ This is strongly suggestive of inadequate protection of family members.

To my mind, the heart of the context for thinking about Social Security is that it substitutes for poor decision making and for missing insurance opportunities (missing perhaps because poor decision making implies low demand). The various shortcomings that are apparent even in the presence of Social Security

⁵ For an overview on 401(k)'s, see Alicia Munnell and Annika Sundén, 2004.

⁶ Deposits into an account each year for 40 years are present in the account for roughly 20 years on average.

⁷ If they existed, one could hedge some of earnings risk through traded indices of wages in different regions and industries. One could try to approximate that through trading in assets correlated with the indices.

⁸ Of course this is dependent on the relative measures of need of singles and couples.

would be more severe in the absence of a program. These different shortcomings in preparation for retirement relate to different issues—inadequate overall provision for retirement relates to having a mandatory program, inadequate annuitization relates to providing benefits in annuitized form, inadequate protection of family members relates to providing benefits for surviving spouses and young children. I start with the annuitization issue, since there is little overt move to end the mandatory nature of Social Security as a whole. But there are calls for decreasing the role of annuities in Social Security.

But first a word on the Arrow-Debreu framework. I have referred to it above as a way to describe the properties of a Pareto-optimal allocation. I think that when economists quickly consider economic issues outside their own sub-disciplines, they frequently think implicitly in terms of the Arrow-Debreu model with its connection to first-best outcomes (also incorporating overlapping generations for some issues). In contrast, economists thinking about issues in their subdisciplines often share a framework that is more complex and are more inclined to do second-best analyses, which are more directly policy-relevant. As in other subdisciplines, analysts of Social Security are well aware of the issues I have identified, although the issues are not present in all analyses by any means. The Arrow-Debreu model tends to start our thinking in terms of the standard, fully rational model of individual decision making and in terms of a complete set of markets. That is a reasonable place to start as long as it is not also the end of modeling and thinking. For example, simulations of Social Security reforms that assume an overlapping generations model with fully rational lifetime utility maximization should not be taken as the whole story for Social Security policy-making. It is inadequate and potentially misleading to study the effects of Social Security in models in which there is no particular reason for Social Security to exist in the first place. This would be akin to treating Pigouvian taxation to correct externalities as distortive by ignoring the externalities.

Interest in the description of behavior that deviates from that in the Arrow-Debreu model has grown enormously lately. Long before behavioral economics became a hot topic, public policies reflected recognition that the model of

homo economicus, while very useful, is not a fully adequate basis for the design of all policies. For example, federal legislation introduced a “cooling-off” period during which contracts with door-to-door salespeople could be cancelled without penalty precisely because of deviations from *homo economicus*. And social security discussions have long recognized inadequate savings for retirement by many workers and inadequate annuitization by most. In addition, social security systems have been concerned about protection of the family and not just the worker.⁹ Also possibly relevant, but not much studied, is whether significant numbers of workers retire too soon for their own good. These issues of poor choices in the presence of available opportunities are in addition to insurance market limitations that come from market incompleteness and from asymmetric information.

Inadequate attention to the future in general and its stochastic structure in particular implies some form of time inconsistency. Normative criteria for evaluating institutions become more complicated without time consistency throughout the population. Insofar as individuals are not time-consistent, it seems essential to do normative evaluations in terms of shorter periods (e.g., years) as well as in terms of lifetimes. We care about actual consumption levels as well as the levels of lifetime resources.

This requires more than just a positive theory of how people determine consumption but also normative criteria for evaluating consumption at different times. The vocabulary of someone being different selves at different times is suggestive, although I am concerned that taking it too literally, failing to recognize the tight links between the different selves who are the same person at somewhat different ages, is failing to address adequately the underlying issues.¹⁰ And

⁹ The growth of two-career families has altered the nature of this concern and presumably the most sensible design for the system, but has not made the problem go away. There is a tension in social security systems, just as there is in income taxation, between treatment of individuals and treatment of the family. Diversity in the way resources are allocated within different families affects the evaluation of different benefit designs.

¹⁰ More generally, there is a difference between taking a model literally and taking it seriously—which involves learning from models in order to think about a reality that is more complex than is captured in any model—indeed that is

since the political process is not equivalent to a consistent approach to policy over time (which, it seems to me, is an essential property of democracy given divergent preferences and views), we must consider issues from multiple perspectives.¹¹

An education built around the Arrow-Debreu model may lead to overvaluing the fundamental welfare theorem. The wonderful properties of competitive equilibrium in certain unrealistic circumstances lead the profession to be very aware of distortions that prevent first-best outcomes. But some distortions are associated with redistribution and with easing other deviations from first-best rules. Stressing the distortions caused by government policies and not giving equal weight to the redistribution and insurance and revenue generation accomplished by these policies, effectively doing partial first-best thinking rather than complete second-best thinking, can lead to unbalanced inferences about policies.¹²

II. Annuitization

Some mandate for retirement saving is not particularly controversial among policy-oriented economists, so I begin with mandatory annuitization. First, let us consider the point of payments that are conditional on being alive. With some saving for retirement (over and above precautionary balances) a worker can learn of rates of return (and risks) available in the market for investing for different lengths of time (that is, including an illiquidity premium). Anyone investing for some period of time (for example, bank certificates of deposit, insurance contracts, mutual funds with an early withdrawal penalty, direct loans) could wonder how much more might be paid if the investor were still alive provided there was no payment at all if the investor were no longer alive. With a

noticeable probability of the investor's dying before reaching the end of the contract period and little cost for checking whether the investor is still alive, it would be worthwhile for a borrower (bank, insurance company, mutual fund, or direct borrower) to offer some additional payment in return for being freed from payment in the event of the death of the investor.¹³

This is the essence of an annuity and the essence of why for an investor with no interest in bequests and a tolerance for some illiquidity, an annuitized asset dominates the same asset without an annuity feature. This is how the Arrow-Debreu model works when markets are complete—the gain from annuitization can be thought of as a lowering of the price of future consumption by forgoing deliveries after one's death. The formal argument for the dominance of annuitization was made by Menahem Yaari (1965) in the context of a conventional annuity that guarantees payments over the rest of one's life. But the argument is much broader than that (Thomas Davidoff et al., 2003). Moreover, simulations show a sizable quantitative importance of annuity opportunities.

People do care about their children. But, a bequest motive does not eliminate the advantage of some annuitization. With a bequest motive and complete Arrow-Debreu markets, one would determine how much of one's lifetime budget constraint to give away and when to give it (e.g., when children reach some age). It would seem very odd to prefer to have one's children receive an amount in present value that was conditional on how long one lived (even if one did not want to make a transfer before dying). So, one would still use annuities, the purchase of commodities conditional on being alive, for all of one's own planned consumption. That is, having a bequest motive is not a basis for doing no annuitization in a complete market setting—unless one was roughly risk neutral about both the amount of bequest and its timing. Without complete markets, a willingness to invest in illiquid assets for future consumption leads to the same advantage for some annuitization.

what a model is all about. On this subject, see Alfred Marshall (1948, p. 366).

¹¹ That is, thinking only in terms of lifetime utilities and sustained government policy rules seems to me inadequate.

¹² For example, when Congress removed the retirement test between the age of full benefits and age 70, some wanted to remove the test for all those over age 62. Noting only that the test discourages work, without noting its effects on the timing and size of benefit receipt would be an example of such partial first-best thinking.

¹³ This is the way a defined benefit system works for workers without any dependents eligible for benefits. The risk could be shifted from the borrower to the set of investors by distributing a fraction of the amount not paid to deceased investors to surviving investors. This is the way that CREF annuities work.

Despite the advantages of annuities, we see only a small fraction of people doing voluntary annuitization.¹⁴ Furthermore, those who do annuitize make very odd choices. They buy nominal annuities. There is wide popularity of what are called guarantees—continued payments after death up to some limit.¹⁵ Such guarantees undo some of the underlying annuitization, and are a relatively expensive form of holding non-annuitized wealth (given the relative administrative costs on annuities and other accounts). They represent an increase in the riskiness of one's bequest, not a decrease. That is, an annuity without a guarantee costs less, allowing one to leave one's heirs a determinate amount in present value, rather than a random amount depending on the date of death.¹⁶ More generally, many features of insurance markets are hard to reconcile with sensible decisions by households and the equilibrium industry response we would expect in the presence of sensible demands. The extremely limited options available for annuitization seem to reflect the natural response of the supply of insurance to the nature of demand.¹⁷

Some have tried to explain this limited use of annuities by the degree of annuitization that already exists in government programs. But voluntary annuitization, while present for centuries before the creation of these programs, was not extensive in the population and is unlikely to become extensive if the programs were removed. Asymmetric information is another candidate for explaining this situation, and it does cause an adverse selection effect on pricing that

would discourage some individuals from annuitizing. While large systematic differences in life expectancy do exist, much of the difference is readily attributed to easily measured factors, so insurance companies could do more to overcome this problem, given the potential for large gains to the insured.¹⁸ In the United Kingdom, there is a sizable market for individual purchase of annuities because of a large tax incentive for their purchase from assets in tax-favored individual retirement accounts. In the presence of this demand, suppliers are offering annuities with better prices for those with "impaired lives." We do not see this risk classification in the United States, presumably because there is not a ready market in which firms could take advantage of selection by risk classification and better pricing since so few households purchase annuities on an individual (nongroup) basis. So, adverse selection alone can not explain the low level of annuitization that is present.¹⁹

I believe the major issue behind this pattern of insurance demand is the failure of many to understand the advantages of annuitization. This plausibly relates to the failure of much of the population to understand the properties of stochastic variables, as has been documented by cognitive psychologists. It is to be expected that the set of insurance products that are marketed will reflect the shortcomings of consumer understanding—it is very expensive to try to sell a product the virtues of which potential customers do not understand. I think that without Social Security, inadequate annuitization would be even more widespread than inadequate savings.

In any event, social security systems in advanced countries typically provide benefits as annuities, annuities that are generally indexed to prices or wages (or a combination). This is a simple application of the view that in a mandatory program, individuals should be given what we think they would want if they were well-

¹⁴ There is a thriving market in what are called variable annuities, but these are tax-favored investment vehicles with a bit of insurance included (to get the favorable tax treatment) and an option to purchase a genuine annuity, an option that appears to be rarely taken. With the recent addition of a lump-sum option in many defined benefit plans, many workers may be forgoing an annuity, although it is difficult to tell since some may simply be waiting to annuitize later.

¹⁵ Among TIAA-CREF annuitants roughly three-quarters choose some guarantee period (John Ameriks, 2002).

¹⁶ Guarantees may play a role in addressing adverse selection, but that is, itself, a reflection of poor functioning in this market relative to ideals.

¹⁷ Although markets provide both term life insurance and whole-life contracts, the only annuities in the market are for payments over the rest of one's life, from the date of the first payment.

¹⁸ Antidiscrimination rules do limit the variables that insurance companies can recognize in pricing. But more could readily be done with allowable categories, such as smoking, type of job, earnings level. The adaptations of the life insurance market to the presence of adverse selection are suggestive that adaptations would occur if the market for annuities were of a comparable size to that for life insurance.

¹⁹ Note that there is considerable risk classification for life insurance.

informed and well-educated. The presence of mandatory annuitization does not prevent bequests, although it raises the cost and requires action to do so. Those surviving to the start of their benefits and with sufficient life expectancy can use part of their monthly Social Security benefits to finance a long-term life insurance contract, thereby providing a bequest with an explicit choice of the relationship between the real size of the bequest and the date of death. This action contrasts with simply leaving unspent funds to one's heirs, a strategy that leaves an amount dependent on the history of consumption relative to the income earned on assets.²⁰

In other words, the government's choice between providing retirement benefits as annuities or as lump sums can be considered as a choice of a default, one which most individuals could reverse—by purchasing life insurance if provided an annuity or purchasing an annuity if offered a lump sum (B. Douglas Bernheim, 1991). Reversing the government choice, though, takes time, thought, and effort and it has a cost. That is, the government provides annuitization at a far lower cost than does the private market. The absence of selling costs (other than equivalent information provision) and economies of scale contribute to this advantage. Administrative costs of Social Security are less than 1 percent of annual expenditures, and a great deal of that is due to the disability program, which is naturally more expensive to run. In contrast, privately provided insurance has higher costs—life insurance company accounting generally recognizes over 10 percent of premia used for administrative costs and profit. The private market is more expensive and does not do a better job of delivering annuity products that people need.²¹ This is one reason to

have government provision rather than a mandate to purchase an annuity in the private market.²² As with many other settings, we expect individuals to undo little of what is provided.²³ So it makes sense to offer what we think people might sensibly want. Moreover, the wider functioning of the life insurance market than the annuities market suggests a further advantage to using substantial annuitization as the default.

A mandatory retirement income program requires a choice of the form of benefit and it is hard to think of a basis for choosing the form which is other than what makes sense for the bulk of the population. It seems to me that this is an annuity in some form.²⁴

A. Lifetime Income Distribution

Mandated annuitization affects lifetime income distribution.²⁵ Suppose one were

nature of the cash-in, cash-out annuity product also suggests that we are not missing cost improvements that would otherwise come. The small size of replacement provided by Social Security leaves lots of room for such private developments if they did represent a significant opportunity.

²² With mandated private purchase, if allowed, we would get risk classification and separate pricing, as has happened in the United Kingdom. Separate risk classification has advantages and disadvantages, and it is unclear whether it would be better.

²³ While many of the elderly have life insurance policies, these appear to arise mostly from coverage for funeral arrangements, small old policies that were not terminated, and tax avoidance among the wealthy, rather than from a conscious attempt to undo annuitization.

²⁴ Mandatory annuitization in a social security program raises the interesting question of how monthly benefit should vary over time—with prices, wages, and possibly other variables such as rates of return. Relevant for this issue are the age structure of optimized expenditures, the relative importance of both real and relative consumption, and the allocation of risk bearing between the elderly and the rest of the population. Currently, benefits in force are increased for inflation as measured by the CPI. While this is a reasonable solution, I suspect it would be better, on a revenue neutral basis, to have lower initial benefits that then grew faster (for example as a weighted average of prices and wages). This would help the longer-lived more than the shorter-lived, but the effect on expected lifetime income distribution could be partially adjusted by changing the benefit formula. But this issue has not received detailed analysis.

²⁵ A full analysis of the income distribution effects of Social Security should consider the disability program along with the retirement income program since there is a negative correlation between life expectancy and the likelihood of collecting disability benefits and dying young enough to leave children who collect young survivor benefits.

²⁰ The timing of the purchase of both life insurance and annuities is important for obtaining insurance. Waiting to insure passes up insurance opportunities. The effects of waiting depend on the extent of risk classification in the pricing of insurance.

²¹ With any commodity, selling costs could be avoided by mandating payment for government provision. What distinguishes retirement annuities from most other products are the similarity of needs of different workers (compared with the diversity of tastes for different commodities) and the reasons why a retirement income mandate is needed. These reasons suggest that public provision is not blocking an otherwise creative dynamic product development. The

comparing two mandatory programs, one with lump-sum payments and one with annuities. This comparison would be easy if individual choices between annuitization and nonannuitization were unaffected by the government choice. Then one would simply compare the implicit price of the trade-off between annuities and lump sums in the alternative mandatory programs with the explicit price at which individuals could make transactions. For example, if everyone annuitized and the market had a single price for all annuities, then we would compare the price implicit in the comparison of the programs with the actual uniform price. In this case, we would find mandatory annuitization attractive because the government would be likely to have a better price than the market.²⁶ Conversely, if everyone would purchase life insurance to undo a mandatory annuity (and rates were uniform), then we would find mandated annuities unattractive since the private market price for life insurance would likely be larger than the implicit price if the government switched from annuities to lump sums.

The story becomes a little more complicated if we assume that everyone annuitizes and the market would offer different prices to different people. This might happen with a mandate to purchase annuities in the private market if the market had some degree of price diversity by risk class.²⁷ Then, in addition to the difference from the average price with and without the government annuity, we would note the differences in prices for different people. Relative to annuities priced differently for different groups, the uniform annuitization implicit in the mandated annuitization would favor those with longer expected lives—women relative to men, male high earners relative to male low earners, female high earners relative to female low earners. A progressive benefit formula can be

used to offset the systematic variation in life expectancy with earnings within genders.

For this or any income distribution comparison, we must have a counterfactual, preferably a plausible one. Without a mandate, the relevant counterfactual is that approximately no one would annuitize. Pretty much everyone would lose the insurance gains from annuitization.²⁸ We can compare the mandate with this counterfactual in two steps—first the value of annuitization assuming actuarially fair pricing and then the difference, described above, between fair and uniform pricing. Since those groups with shorter life expectancies have more to gain from fair annuitization [assuming CRRA preferences in the usual range and realistic mortality rates (Jeffrey Brown, 2003)] this counterfactual shows much less diversity in the utility value of annuitization than the previous comparison.²⁹ Indeed, Brown finds that the utility value of annuitization (relative to wealth) is similar for groups divided by gender, race, and education. Thus the differences in expected payments from different life expectancies have less distributional impact in utility terms than in expected payment calculations.³⁰

²⁸ I ignore the role of access to minimum income guarantees (SSI).

²⁹ Someone with a higher probability of dying would find a larger decrease in the price of consumption when going from unconditional purchase to purchase conditional on being alive with fair pricing. Without annuitization, someone with a higher probability of dying would generally consume less in later years, *ceteris paribus*, and so have less consumption on which to receive a price decrease. Given the preference structure in Brown, the net result of these two effects is that those with shorter lives gain more from fair annuitization, tending to offset the redistribution from a change from fair to uniform annuity pricing. Interpretation of the Brown analysis of the total impact of annuitization is aided by the analysis in Bernheim (1987) of the valuation of marginal annuitization relative to life expectancy and liquidity constraints.

³⁰ Implicitly this income distribution discussion has assumed rational lifetime consumption allocation—the only behavioral element being an unexplained, and unexplainable (on rational grounds), failure to purchase annuities. Similarly, the simulations showing the value of annuitization assume optimal consumption paths both with and without annuitization. Any full normative evaluation of annuitization should reflect the fact that those living longer after retirement will have a larger marginal utility of consumption for any given wealth for retirement consumption.

²⁶ W. H. Beveridge (1942) argued that in the United Kingdom the government systematically did better than private insurance markets.

²⁷ Annuity pricing that varies with stochastic health outcomes implies a risk of the classification to which one will belong. With annuitization done at a single time in life, the degree of risk classification involves a tension between providing more insurance and providing more accurate labor market incentives.

B. *Labor Incentives*

The implicitly uniform-price annuitization in Social Security also affects labor market incentives. The use of uniform annuity pricing (overall or within still heterogeneous risk classes) violates the conditions for first-best optimization. Compared to first-best pricing, the decision that would be distorted is that of labor supply. If annuity pricing is breakeven, then some are being taxed on work while others are being subsidized compared to a system where annuities are priced for individual life expectancies. An alternative counterfactual would be a failure to annuitize at all. Without annuitization, we would have more accurate labor market incentives person-by-person, but earnings would finance less satisfactory consumption trajectories. We would fail to insure not only life expectancy realizations but also changes in life expectancy as information accrues. That is, even unfair annuities can raise individual welfare if the alternative is no annuities.

I conclude that having a mandated retirement income program provide its benefits as annuities is sensible.

III. **Workers and Families: Young Child, Spouse, and Survivor Benefits**

Social Security provides more than just retirement benefits for workers. It provides benefits for disabled workers and their families, for young children of a deceased worker, and for elderly spouses and surviving spouses. In addition, a divorced spouse may be eligible for the same benefits as a spouse if the marriage lasted at least ten years.³¹ Benefits other than worker benefits are referred to as auxiliary benefits. These benefits are subject to a maximum rule—a beneficiary receives the largest benefit he or she is eligible for—with no increment for also being eligible for a smaller benefit. That is, if someone has worked at least ten years, on retirement he or she is eligible for a retired worker benefit. He or she is also eligible for a spouse benefit if married to a retired worker beneficiary. But the total amount of benefit is equal to the maximum of the two benefits. Sim-

ilarly, someone eligible for a worker benefit and also eligible for a survivor benefit receives only the larger benefit.³² A central design feature is that these auxiliary benefits are not paid for on an individual basis—workers with the same earnings history receive the same retired worker benefits whether or not they have family members or ex-spouses collecting auxiliary benefits.

Auxiliary benefits raise four questions. Does it make sense to mandate benefits for family members and ex-spouses? Does it make sense to base benefits on a maximum rule? Does it make sense to finance all of the auxiliary benefits from the program as a whole rather than in part or in full from the benefits of the retired worker? Are the details of benefit determination rules as well-designed as might be?

Let me start with the first, most basic question. It makes sense to provide auxiliary benefits since studies suggest that significant numbers of workers do not insure their lives adequately and would not make good choices between single- and joint-life annuities. More generally we are learning more about the ways in which the allocation of resources within the family does not conform to a single maximization with a single budget constraint. Since the government cares about the different family members (and not just the worker), direct allocations to family members matter since they will change the allocation of resources within the family. Protecting family members is a role governments have recognized for centuries.

The other questions are more complex and need more detailed analyses. Two issues are central here. These are the positive and normative issues of how consumption is actually allocated within families and how to combine evaluations and rules that affect both individuals and families. Research on the determination of allocations within the family is still in an early stage of development. And normative analysis has not progressed much beyond identification of the dilemma in recognizing both

³¹ There is a family maximum that proportionally reduces all benefits except those of the worker if it binds.

³² If the spouse or survivor benefit is larger, the person is referred to as having dual eligibility. In 2002, 38 percent of elderly women received only a worker benefit, 34 percent received only an auxiliary benefit and 28 percent were “dually entitled.” These fractions are expected to change as more and more women with substantial careers reach retirement age.

individuals and families.³³ So my answers here are speculative and primarily meant to identify where research might lead us to policy improvements.

Offhand, the maximum rule does not provide labor incentives well (incentives are stronger for workers who have spouses likely to collect auxiliary benefits since two benefits are increased by more earnings, and weaker for workers who are likely to receive a larger spouse benefit since further earnings by someone receiving a spouse benefit do not increase that benefit). A similar (but less extreme) issue arises with income taxation of a lower-earning spouse.

Offhand, the cost of auxiliary benefits should be shared between a worker and the program as a whole. The benefit formula is progressive, with a higher replacement rate for lower earners, reflecting differences in retirement needs. As part of responding to needs, it seems right to recognize dependents in determining need and so benefits. But the current rule is not the only way to do that. Some provision of auxiliary benefits for children from the general program makes sense, in keeping with our generalized support of children in education; some provision for spouses is relevant in the role of progressivity—two can not live as cheaply as one. But the current system has gone too far and I share in the criticism that too much is given to the nonworking spouses of high earners.³⁴ Using system resources to finance large transfers to those in the upper tier of the earnings distribution offsets too much of the progressivity in other portions of the system. Designing a different system would be politically sensitive and complex and would need detailed analysis.

The determination of survivor benefits has also received considerable criticism. Recogniz-

³³ If one wants to do a normative analysis solely on the basis of individual experience, one needs to consider how resource allocation within the family is affected by the rules determining benefits and the impact of benefits on marriage rates. Such an analysis could consider the effects of extending auxiliary benefits to all long-term relationships, including same-sex marriages.

³⁴ Scaling back the spousal benefit for spouses of high earners (for example by a cap), or a more general overhaul of auxiliary benefits is likely to meet considerable political resistance. This suggests not tackling this issue in reform plans hoping for an early restoration of actuarial balance. In addition, a general overhaul should be preceded by considerable further analysis.

ing the role of couples in sharing resources at least partially, it makes sense to relate the benefits of an elderly survivor to the benefits that had been received by the couple—a survivor replacement rate. Currently, survivor replacement rates vary with the past earnings of husband and wife, usually, but not always, between one-half and two-thirds.³⁵ A uniform survivor replacement rate seems more likely to approximate relative needs than the current system. The much higher poverty rate of widows than of couples, noted above, suggests a higher survivor replacement rate is needed, with three-quarters having been suggested by some analysts.³⁶ The change to a uniform and higher survivor replacement rate could be financed out of a suitable mix of the total resources of the program and the benefits of the couple while both are alive.

The current recognition of divorce is to allow benefits for unremarried divorced spouses and divorced surviving spouses after at least ten years of marriage. Since there is a family maximum, some of these benefits are paid by the system as a whole and some out of the other auxiliary benefits. The adaptation of the system for the growth in divorce seems to me a major issue for research. I do not know if we can design a system that would be better, recognizing both labor market incentives and income distribution issues, or if such a design could survive political hurdles.³⁷ But it is worth thinking hard about.

³⁵ It is common to cite the range of one-half to two-thirds for the survivor's replacement rate, ignoring actuarial adjustments. But there are cases above this range once we include adjustments for the ages at which the benefits are claimed.

³⁶ Among the TIAA-CREF annuitants who choose a joint-life annuity from the three available options, roughly 70 percent choose a full benefit to the survivor, nearly 20 percent choose two-thirds, and the rest choose one-half. Roughly 70 percent of men and 30 percent of women choose joint-life annuities (Ameriks, 2002).

³⁷ Research hurdles come from combining concerns about individuals and families. Incentives for retirement depend on benefits relative to individual earnings, while need reflects family incomes. Political hurdles come from the diversity of views about the structure of benefits. Some like the discouragement of labor force participation by those with children, others prefer not to subsidize that activity. Divorced women are among the most vulnerable beneficiaries. With the benefits for divorcees with limited earnings tied to the benefits for spouses, reducing spouse benefits for high earners affects both well-off and vulnerable groups. Finding a way to satisfy diverse constituencies will not be easy, as is shown by repeated groups looking at this issue.

In sum, mandating benefits for the families of workers is important, along with mandating savings and mandating annuitization—the inclusion of family benefits in Social Security makes sense. There is good reason to think that the current rules can be improved, but research difficulties and political hurdles will need to be overcome if we are to make improvements.

IV. Income Distribution, Insurance, and Labor Supply³⁸

In determining retirement benefits, Social Security first averages the best 35 wage-indexed annual earnings,³⁹ then it uses a progressive benefit formula to determine what real benefits would be if first claimed at the age for full benefits (commonly, if somewhat misleadingly, called the normal retirement age),⁴⁰ and then it adjusts benefits for the age at which they start. Moreover, between age 62 (the earliest age at which retirement benefits can be claimed) and the age for full benefits (which is in transition from 65 to 67), benefits are only paid if earnings are low enough, referred to as an earnings or retirement test. Each of these steps in determining benefits affects income distribution, insurance, and labor supply. I will skip over implications of using 35 years (as opposed to more or fewer years or all of lifetime earnings subject to tax)⁴¹ and of using a wage index to weight the earnings in different years in determining benefits (as opposed to using an interest rate)⁴² and concentrate on the effects

of a progressive benefit formula and a retirement test.⁴³

Consider the stochastic process of earnings opportunities. Individual workers face considerable risks that are only partially correlated with the economywide average earnings used in indexing. Wages move differently by industry and firm and region and some individuals have career opportunities strongly affected by industry and firm and region developments. We do not have trading in the type of indexes Robert J. Shiller (1993) has proposed in order to give workers the ability to hedge these aspects of their risks.⁴⁴ Even if we managed to have trading in such indices, it is beyond credibility that most workers would take appropriate advantage of these opportunities. When many workers can not sort out the basics of portfolio diversification in their 401(k)s, there is no reason to anticipate successful execution of far more complex financial strategies. By having replacement rates that are higher for lower levels of lifetime earnings, a social security system that

a cohort in each case. Using a wage index gives less weight to the earlier years and more weight to the later years than does use of an interest rate. Thus, use of a wage index would be implicit taxation on younger workers and implicit subsidization of older workers and would be a distortion in a first-best world. However, the annual income tax is progressive, so that an upward-sloping age-real earnings profile implies that on average older workers have higher marginal income tax rates than younger workers. Thus, the sum of the marginal income tax plus the implicit Social Security tax is smoother across ages with a wage index than with weighting by interest rate. Also, preferences are not intertemporally additive. The standard of living to which retirees have become accustomed is more affected by earnings later in life than earlier in life. While the effect of introducing a standard-of-living effect into annual utility has been explored in simulations of the value of annuities (Davidoff et al., 2003), no similar analysis has been done for the weighting given earnings in different years.

⁴³ I do not discuss the issues behind the choice of 62 as the earliest age of eligibility for retirement benefits (EEA). Increasing the EEA helps those who would otherwise retire too early for their own good and hurts those who are right in their early retirement decision and are hurt by the illiquidity from benefit nonavailability until the EEA. Measuring the size of the two groups would be very hard and only a little has been done in identifying people who are affected. Increasing the EEA would have little effect on long-run Social Security financing as explained below. Similarly, it would be hard to design a good method for automatically indexing the EEA.

⁴⁴ Robert C. Merton (1983) has examined the role of Social Security in sharing aggregate earnings risks more widely.

³⁸ For a discussion of the links among tax theory, incomplete market theory, and social security, see Diamond (2002).

³⁹ Earnings after age 60 enter in nominal terms, not in indexed form. This should be changed to have labor incentives not vary this way with inflation.

⁴⁰ Such progressivity is missing in Europe, where income distribution issues are addressed more fully in other parts of retirement income provision.

⁴¹ Depending on the nature of the underlying stochastic process of wage rates, both underweighting early years (relative to the use of interest rates) and not counting some low years may or may not help with insuring lifetime earnings—this is not an area that has received much research attention.

⁴² We can contrast the earnings incentives between benefit formulas that accumulate earnings with a wage index and those that accumulate with a (presumably higher on average) interest rate. This is most readily done in a breakeven comparison, assuming the same level of resources for

makes benefits a progressive function of lifetime earnings offers insurance about lifetime earnings that is not available in the market.

If the taxes and benefits for a cohort broke even in present value terms, the use of a progressive benefit formula would imply that the labor supplies of lower earners were being subsidized and those of higher earners were being taxed.⁴⁵ This is the familiar pattern with insurance with asymmetric information—a combination of insurance and incentives neither of which satisfy the conditions for first-best optimization. This effect of progressivity is in addition to the effects from annuity pricing discussed above. Some of the effects of annuitization and progressivity would be offsetting—those with higher earnings of each gender tend to live longer—and some would be compounding—women on average have lower earnings and longer lives. That taxes and benefits do not break even on a cohort basis is discussed in the next section.

The progressivity in the benefit formula uses taxes that distort labor supply in order to redistribute income and provide insurance. The progressive annual income tax also redistributes income, provides insurance against earnings uncertainty, and distorts labor supply. Since these two institutions work on different tax bases and provide payments at different times, there is room for each of them to contribute despite the presence of the other. Annual income taxation recognizes short-term needs, coming from borrowing constraints and from behavior that is not time-consistent. It also recognizes capital income as part of determining tax rates. *Ex post*, all of one's Social Security taxable earnings (in the best 35 years) contributed to benefits in a way that varies with age but not with the level of annual earnings, given lifetime earnings. This avoids the distortions coming from having different marginal tax rates in different years as a function of annual earnings, or annual capital income. The use of a lifetime measure also separates out issues of lifetime earnings from the age-earnings profile in doing redistribu-

tion.⁴⁶ While both annual income taxation and lifetime social security have received analyses of the trade-off among redistribution, insurance, and distortions, there has not been much work considering the simultaneous use of both institutions.

A. Retirement Test

For a mandate to save for later consumption to have bite, workers can not be allowed to claim benefits whenever they want, including immediately. To claim Social Security retirement benefits, a worker must be at least 62. The system could simply start paying benefits at age 62. Instead, between age 62 and the age for full benefits, workers can only start receiving benefits if their current earnings are low enough, corresponding to full or partial retirement for many workers.⁴⁷ Any delay in the start of benefits increases their monthly amount, tending to counterbalance the delay in the start of benefits. The impact of this retirement test on labor market incentives is in addition to effects discussed above that apply to each year of labor supply. That is, the effect of Social Security on incentives for continued work past age 62 has two parts. One is the effect of a delay in the start of benefits together with their later increase as a consequence of the delay in their start.⁴⁸ The

⁴⁶ For example, if everyone had the same age-earnings profile, Social Security would do no redistribution within a cohort, while annual income taxes would subject each person to earnings subsidies when younger and taxes when older.

⁴⁷ Benefits are paid to workers younger than the age for full benefits if earnings are below the exempt amount, which equals \$11,640 in 2004. Earnings above this amount result in a 50-percent reduction in benefits, until benefits reach zero. Rules are different for the year in which the age for full benefits is reached. After reaching the age for full benefits, benefits may be claimed whatever the ongoing level of earnings. A worker can receive a larger benefit by delaying the start of benefits up to age 70.

⁴⁸ The start of benefits can be delayed even if the worker retires. For a worker without liquidity constraints, labor supply is not encouraged by a net subsidy from delay (in the case of a long expected life) since delay is available anyway. But work is discouraged for those with shorter life expectancies. While some eligible workers do not claim benefits right away, overwhelmingly, retired workers do claim fairly quickly. For those who would claim as soon as they stopped working, work is encouraged by a larger increase in benefits from delay as a result of a longer expected life.

⁴⁵ This resembles an EITC being financed by a positive income tax. Unlike the EITC, which has a region of high marginal taxes as the subsidies are phased out, Social Security has a monotonic transition from marginal subsidies to marginal taxes.

second is the extent to which additional work, and so additional payroll taxes, increase the measure of lifetime earnings and so add to benefits.

For an average worker at ages 62 and 63, Social Security had a roughly zero marginal tax for the average worker when the age for full benefits was 65.⁴⁹ With the increase in the age for full benefits there will be a small tax at these ages. While implicit taxes used to be much larger above the age for full benefits, the retirement test has been eliminated for those ages. With differences in life expectancy, a zero tax on an average worker implies that some workers are taxed and some are subsidized by the presence of the retirement test.⁵⁰

The retirement test has two effects. One is to raise (delayed) monthly benefits for those continuing to work. To the extent that a worker would have consumed out of benefits received while still working, the delay in the start of benefits raises later consumption (for both the worker and possibly a surviving spouse) since more is saved. This is advantageous to the extent that consumption falls too much after retirement.⁵¹ The combination of a delay and increase in benefits is also redistribution across workers based on life expectancy along the lines discussed above. On the other hand, benefit ineligibility while continuing to work discourages work for those not fully valuing their increased later benefits and those with shorter life expectancy. Empirical estimates find that the overall labor supply effect is modest, suggesting that the increase in monthly benefits effect is more important. The retirement test also helps

with the risk in earnings trajectories that comes from how opportunities to earn (and disutilities) develop toward the end of a career. The retirement test addresses that risk to the extent that there is taxation on continued work and those continuing to work are less needy on average than those who stop working earlier.⁵² Thus I conclude that the retirement test does distort labor supply, but that distortion is more than offset by the gains from improved lifetime consumption allocations and increased insurance.

Limiting the range of ages at which the retirement test applies makes sense. Otherwise some of those working to very advanced ages would have replacement rates above 100 percent and, if liquidity-constrained, would prefer to have part of benefits while still working. Currently the age for the end of the retirement test is the age for full benefits. I am not aware of any analysis of the optimal choice of an age for the end of the retirement test.

B. Labor Supply at Younger Ages

I have focused on the retirement decision since elasticities here are larger than those with earlier labor supply decisions. But younger workers pay payroll taxes and anticipate an increase in benefits once they retire as a consequence of the earnings that were subject to tax. The effect on labor supply is relevant for choosing the size of a mandatory retirement income system. This incentive depends on the perceived link between taxed earnings and retirement (and disability) benefits. While those nearing retirement age often gather information on the workings of the system and seek advice on the advantages of different timing of retirement, younger workers are not well informed.⁵³ Some

⁴⁹ Courtney Coile and Jonathan Gruber (2001). The marginal tax reflects the loss of a year of benefits, the increase in benefits thereafter, the payment of the payroll tax, and the increase in benefits from an increase in AIME, since earnings in a late year are likely to be among the top 35. For example, of the Health and Retirement Survey sample of those born in 1931, over one-third of men and four-fifths of women had fewer than 35 years of positive earnings when entering the year in which they turned 61. Without a tax on continued work, increasing the age at which benefits can first be claimed, without other changes, would not save money for Social Security on a permanent basis.

⁵⁰ Of course, these earnings are also subject to the annual income tax.

⁵¹ Also valuable is that the higher monthly benefits come as an annuity while savings from benefits would not provide this insurance.

⁵² I note that in a model with homogeneous life expectancy, if we have fully rational workers and if the increase in benefits implies an implicit tax, there is an increase in insurance insofar as early retirement is a consequence of an adverse realization of opportunities. This is a familiar optimal insurance result—that one taxes observable variables in the states where they signal a lower marginal utility of consumption.

⁵³ Information is also supplied in annual statements which give individual benefit levels for different retirement ages. Someone anticipating no benefits at all might see no link (although inconsistent anticipations about the future are common). Presumably that would change with a reform that was widely perceived as restoring long-run sustainability.

simulations have assumed that younger workers perceive no increase in future benefits as a consequence of additional earnings. This leads to a big boost in apparent efficiency from a switch to individual accounts if it is also assumed that money going into individual accounts has no implicit tax. Both of these assumptions seem wrong to me.

I believe that there is wide awareness of the existence of some link between earnings and later benefits, although understanding of how the link works is not so wide. Misperception of the link sometimes takes the form of imagining that Social Security is like a corporate defined benefit pension that heavily weights later years. This perception would correspond to an implicit tax at some ages and an implicit subsidy at others, not a full tax at all ages. The extent to which labor supply is affected by concern that there will be no benefits would be greatly modified by any reform that restored actuarial balance, not just one with individual accounts. Insofar as workers have high subjective discount rates, mandating savings in any form affects labor incentives and the exact link between taxes and benefits is of reduced consequence. My sense of a small difference between pension systems in incentives for younger workers is supported by the evidence of quite modest labor market responses in Latin American countries that have introduced individual accounts.

I have now argued for the use of a mandatory retirement income system paying annuitized benefits to workers and their families based on a progressive benefit formula and using a retirement test at some ages but not at others. I turn next to two issues that bear more on reform options, as well as reflecting the history of the system. First I will discuss the redistribution across cohorts and then the use of automatic indexing as well as periodic legislation.

V. Benefits by Cohort

Social Security is often criticized for distorting labor supply and savings. Despite the linking of these two decisions, the issues are very different. I have already noted that mandatory annuitization with uniform pricing distorts labor supplies relative to an idealized alternative, but seems to be a welfare improvement relative to a world with no annuities. And I discussed other labor market issues where Social Security com-

bines incentives with redistribution and insurance. In contrast, the rules of Social Security do not distort savings. That is, Social Security certainly affects savings and so national capital. But the term distortion is usually reserved for an intervention that would prevent Pareto optimality in an economy that would otherwise satisfy the conditions of the Fundamental Welfare Theorem. To examine this meaning of distort (as opposed to merely change) we need to consider the impact of Social Security on the marginal return to private savings (the size of a tax wedge). By itself, Social Security has no effect on the return to marginal savings since benefit levels do not depend on capital income. Social Security does interact with the income tax, but the effect of the existence of Social Security on the income taxation of the return on marginal savings can have either sign for differently situated workers, although it probably includes a wedge on average.⁵⁴

A mandate to pay taxes and receive benefits would affect private savings even if there were no marginal distortion at all. Effects come from the requirement that people pay taxes at levels and times when they might not have saved the same amount. Effects also come from redistribution, both within and across generations, that is, from income effects as opposed to substitution effects. I am not aware of any study of the impact on savings from the progressive benefit formula—the presence of higher benefits relative to taxes for low earners who have a lower

⁵⁴ Perhaps the largest effect comes from the cutoffs below which benefits are not taxed or are taxed at a lower rate. Since the cutoffs are compared with income including capital income, there is an increase in the tax wedge on savings for those who are affected in this way. Another effect comes from the possibility that taxable benefits might increase the marginal tax bracket. But Social Security displaces some private savings. Whatever savings are displaced by Social Security might themselves have affected the marginal tax rates (depending on the tax treatment of displaced savings). Moreover, one needs to consider the effect of Social Security on the income tax rate when the savings are done as well as when the proceeds are received—the employer share of the payroll tax is not part of taxable income for the income tax. Thus, the effect could be positive or negative depending on the level of displaced savings and their income tax treatment. This indirect link is present in many other programs that are not talked about in this way. For example, government support of education raises earnings and so marginal tax rates. The implied increase in savings distortions does not seem to be of consequence.

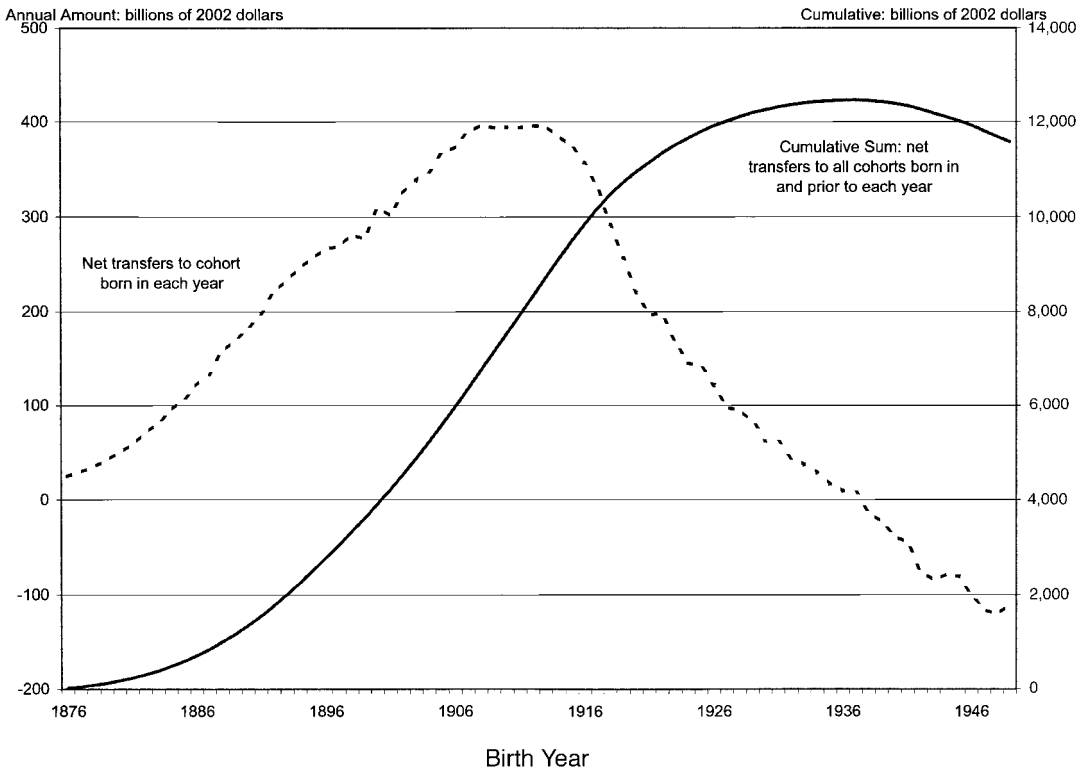


FIGURE 1. NET INTERCOHORT TRANSFERS UNDER SOCIAL SECURITY

Source: Leimer, "Cohort-Specific Measures of Lifetime Net Social Security Transfers," Social Security Administration Office of Research and Statistics Working Paper No. 59, February 1994, updated to present value 2002 dollars.

propensity to save than high earners.⁵⁵ The redistribution across generations has received particular attention and has led to consideration of the impact on national capital.

A. *Transfers by Cohort*

Everyone is aware of the decision to pay earlier cohorts of retirees benefits far larger than could have been financed by the taxes they paid and the interest that could have been earned on them. Figure 1, an updating to 2002 dollars of analysis done by Dean R. Leimer (1994), shows the lifetime transfers by cohort (left scale) and the cumulative net payments by cohort (right scale) for cohorts born through 1949, and so

turning 55 this year.⁵⁶ The aggregate net transfers to these cohorts is roughly \$11.5 trillion.

How much did this early generosity reduce national capital? We have some estimates but they are surely not reliable. A believable time-series econometric study is probably not doable and there is no consensus that one has been done satisfactorily. Another approach would be by simulation. But a credible simulation requires modeling the appropriate underlying behavior—the extent to which different workers would save on their own without such a program. Surely, a simulation with all workers being fully rational lifetime utility maximizers has no credibility. And we would also need to track the effects on national savings from Social

⁵⁵ The difference in propensities to save is also relevant for the impact of any use of payroll tax revenues to lower income tax rates.

⁵⁶ To extend the figure to later cohorts, we would want to consider how actuarial balance is restored, and I do not present such a figure. Leimer assumed phased-in tax increases to balance the present value of taxes and benefits.

Security displacing transfers to the elderly from the government (through the program for the poor elderly—Old Age Assistance, which became SSI) and from individuals (through cash gifts and shared housing).

While the impact on national capital would be an interesting positive question if we could answer it well, it is important to recognize the additional issues needed for a normative analysis. The goal of Social Security's early generosity was to raise the consumption of early cohorts of elderly. Apart from business-cycle effects, higher consumption implies lower savings—implying that lower national capital was required by the goal, not an unintended side effect. A normative evaluation of the impact of the redistribution to early cohorts would consider how much their wages were lower than those of later cohorts and how little they had saved, as well as the return on capital. It would also consider the pattern of transfers within benefited and paying cohorts. However such an analysis would come out—balancing very worthwhile transfers with some less worthwhile ones—most of the transfers are now history.

Given the infinite horizon present value budget constraint of Social Security (in the absence of transfers from general revenues) this early generosity is the cause of lower benefits in the future than could otherwise be afforded. That is, the legacy of the early generosity of Social Security shows up in assets that are not there. If they were present, they would be earning interest that could contribute to paying for benefits. The cumulative curve in Figure 1 gives a sense of the magnitude of the trust fund that is not there because of Social Security's history. But, Figure 1 is by cohort, and so does not show how much larger the trust fund would be today if every cohort had been on a breakeven basis. Although such a calculation is doable, it would not be the best basis for insight into reform options. Rather, that comes from considering the elements likely to constrain reform. Past payments are history and political considerations suggest that it is unlikely that benefits will be directly reduced for those already retired or those nearing retirement, although these benefits might be affected by changes in tax treatment or in the inflation indexing of benefits, changes that would apply to everyone. A partial picture of that constraint would be that cohorts

over 55 would not be affected by reform.⁵⁷ The measure is not exact since cohorts over 55 would be affected by any payroll tax change and slightly younger cohorts are likely to have limited changes in benefits as we phase in any benefit reductions that are part of a reform. An ideal definition of this constraint would conform to a theory of political constraints on reform coming from past generosity. We do not have a full theory, but this gives a reasonable sense of the size of the legacy that needs to be financed from future cohorts.

Peter Orszag and I have referred to the missing assets on this cohort basis as a legacy debt.⁵⁸ Thus the legacy debt is not a debt in the traditional sense of that word, but that term crystallizes the need to allocate the cost of the assets that are not there across cohorts. Spreading the cost of that early generosity across cohorts is inherent in any plan that restores actuarial balance. While only an approximation to the real constraint, the number is roughly \$11.5 trillion (a bit more than one year's GDP). If we were to go to full funding, then this is roughly the cost that would fall on the generations during the buildup to full funding. Alternatively, instead of ever achieving full funding, we can consider a wider allocation of the legacy cost by aiming to preserve a ratio of the legacy cost to taxable payroll. This would parallel the idea of preserving the ratio of the public debt (or the interest on the public debt) to GDP. Spreading the legacy cost over all future cohorts implies less than full funding of Social Security. Without extensive evaluation of its consumption transfers, the effect of Social Security on national capital is not, by itself, a basis for concluding that the system should have been fully funded or should become fully funded.

The baby boomers are much larger than earlier cohorts. The 1983 legislation included payroll tax revenues in excess of current outlays in order to build a trust fund which would then be used to finance the retirement of this very large cohort. That is, taxes were higher early to allow

⁵⁷ In his charge to the Commission to Strengthen Social Security, President Bush included the principle that Social Security reform not affect the benefits of anyone 55 or older (Commission to Strengthen Social Security, 2002).

⁵⁸ This is the same as the "closed group" measure of balance, with the group including everyone 55 and over.

them to be lower later.⁵⁹ Politically, the trust fund is very likely to be used for Social Security purposes in the sense that the constraint on future Social Security expenditures includes the value of the assets in the trust fund. A separate issue is the extent to which the higher payroll taxes since 1983 increased national capital.⁶⁰ This is a source of controversy, with a wide range of presumptions and no ability to settle the question econometrically.⁶¹ I believe that a large part was saved—despite the large federal deficit outside Social Security for the 1980's and early 1990's. In my view, a larger unified deficit, if Social Security had not been in surplus, would not have had a strong effect on tax and spending legislation. Congress had great difficulty in legislating tax and spending changes to lower the deficit. Without the Social Security surplus, a somewhat larger unified deficit would not have changed the basic character of the situation—a deficit widely perceived as being too large and a difficulty in raising taxes and lowering spending. Looking beyond the baby boomers we do not currently perceive a need to single out a cohort that will differ greatly from others and perhaps call for something other than a smooth adjustment of taxes and benefits.

Redistribution across cohorts has not been done in a lump-sum fashion, but through the choice of tax rates and benefit formulas. Thus the redistribution has affected labor supplies as well as savings decisions. In the early days, the generous benefit formulas (in effect or anti-

pated in the future) subsidized labor, just as the lower benefits relative to taxes for younger workers today taxes labor.⁶² This is similar to the role of the progressive benefit formula discussed above. Given the pattern of redistribution by cohort shown in Figure 1, much of redistribution served as an incentive for much of the working life of recipients. This is in contrast with analysis in two-period models where the initial elderly recipients of transfers receive a lump-sum transfer and all later cohorts have implicit taxes on earnings to pay for it. Both the transfers and the taxes have influenced labor supply.

VI. Automatic and Legislated Adjustments to Aggregate Realizations and Risk Allocation

The actuarial projection for the 1983 reform envisioned a buildup of the trust fund, followed by its decline back to the precautionary level of one year's expenditures at the end of the 75-year projection period. It has not worked out that way. Instead of having just enough money for 75 years of expenditures, plus a small trust fund at the end, it is now projected that the trust fund buildup will be sufficient to pay currently scheduled benefits only until 2042. That is, the policy that was designed for a 75-year horizon will, if the projection is correct, cover all of expenditures for only 60 years. By the scale of preparing for long-term outcomes, that does not seem to me to be too bad. Of course one could argue, with hindsight, that Congress should have looked further into the future than 75 years, although it was hard enough to reach agreement on legislation even with that target.

Current discussions have extended the notion of actuarial balance to include "sustainability"—that the ratio of the trust fund to annual expenditures not decline at the end of the horizon. This criterion of sustainable solvency is meant to avoid a repeat of the post-1983 experience where the projected actuarial deficit returned quickly (although the trust fund exhaustion date was distant). Projected deficits returned quickly because of what is called the terminal year problem, or the cliff problem. That is, each year,

⁵⁹ That has shown up in the assets in the trust fund—currently over \$1.5 trillion or roughly 2.8 times annual expenditures. This ratio of trust fund to expenditures is projected to peak at 4.7 in 2016.

⁶⁰ Note that the analysis needs to be done in terms of taxes and benefits, the causes of changes in the trust fund. The question of the impact of the trust fund on national capital requires a distinction among different ways in which trust fund size can be changed.

⁶¹ There has been a relatively short time during which there is a plausible linkage between Social Security and the rest of the budget. Moreover, specific pieces of legislation imply different time shapes of revenue changes and spending changes over subsequent years. Thus, there is no simple link between deficits and lagged deficits (or between unified deficits and Social Security surpluses) that could reliably be discovered by time-series analysis. In particular, it is not credible to believe that econometric analysis could uncover the counterfactual pattern of taxes and spending that would have occurred if the 1983 Social Security legislation had involved lower tax rates.

⁶² The effects of benefit increases of the already retired (through 1972) affect labor supply insofar as they were anticipated earlier. A similar issue arises for benefit increases at different times during a career.

the realized net cash balance of Social Security is added to the trust fund and another year is added at the end of the 75-year horizon. With a constant tax rate and the current benefit formula, the added year is in worse fiscal shape than the average of years before. Indeed of the current 75-year actuarial deficit of 1.9 percent of taxable payroll, a full 1.1 percent is due to the fact that the projection now goes 20 years further into the future than it did in 1983.

The 1983 legislation included future decreases in benefits by increasing the age for full benefits. At the time of the 1983 legislation, there was still a tax rate increase on the books. That was kept and took effect in 1990. Indeed from the initiation of the program in 1935 until 1990, there was always a future tax rate increase on the books. Given the political ease of raising benefits or cutting taxes, and the political difficulty of raising taxes or cutting benefits, having future tax rate increases and future benefit decreases on the books lowers the political cost of preserving balance, since it is easier to legislate future pain than current pain. Avoiding a recurrence of actuarial imbalance a short time after reform requires a substantial trust fund at the end of the projection period, so that it can fall for awhile without triggering imbalance, and/or a change in the time shape of taxes and benefits. A changed time shape can be legislated directly (as we legislated an increase in the full benefit age in 1983 and have legislated future tax increases) or could be expected from the adoption of further automatic adjustments (for example, by including an adjustment for life expectancy).

Before considering the choice between legislated changes and automatic adjustment, let us consider the allocation in a complete-market Arrow-Debreu equilibrium. In the model, outcomes are fully specified. Given subjective beliefs about the probability structure of the states of nature, one can express the value of equilibrium for an individual. Also fully specified is standard modeling of incomplete markets, which replaces complete market auctioneer-announced future allocations by accurate predictions of future market equilibria as repeated trading unfolds. Time-inconsistent individual behavior does not interfere with the ability to describe outcomes in this way, although it will generally interfere with the efficiency properties of equilibrium.

Most social security systems lack the completeness that is needed to specify outcomes solely in terms of underlying economic variables (and the stochastic structure of states of nature). U.S. legislation determines the payroll tax rate for each year into the indefinite future. The level of earnings that are subject to tax each year is automatically indexed—thereby relating taxable earnings to economic outcomes.⁶³ Legislation also sets down the rules for benefit payments in terms of individual earnings histories and price and wage indices. While each part is fully specified, no mechanism ensures that the Social Security budget constraint is satisfied. Thus, there is the expectation that sooner or later something will have to be changed. That is, in order to model future labor and consumption, we need to model future legislative outcomes. This is hard.⁶⁴ In some exercises, the Congressional Budget Office has been instructed by law to ignore some possible future legislation (such as extensions of sunsets of income tax changes). But this is not a satisfactory solution for academic analysts, nor for individuals who are making lifetime plans.

We have a theory using incomplete contracts as part of the theory of the firm. In that theory agents have well-defined property rights and well-specified behaviors that determine the outcomes not covered by the contracts. With incomplete legislation, the future legislative process plays a key role in determining outcomes that are incompletely specified.⁶⁵ Analyzing an equilibrium that includes a legislative process is difficult—requiring modeling the interaction of the personal preferences of elected officials with their concerns about reelection, as well as election outcomes (R. Douglas Arnold, 1990). It is not that this is unknowable in principle, but that we are a long way from a genu-

⁶³ Some of the income tax revenue from the taxation of benefits goes to Social Security as well. This revenue is dependent on future income tax rates.

⁶⁴ In the list of reasons why members of the Panel on Privatization of Social Security of the National Academy of Social Insurance disagreed on the advantages of individual accounts, a central element was the divergence of views on the political implications of accounts—particularly the sustainability both of rules for the accounts and of the structure of traditional benefits (Diamond, 1999).

⁶⁵ The legislative process can also change what is completely specified, but at least we can analyze what happens if there are not any changes.

inely usable, empirically validated theory and we are studying a process that generates very limited data relating outcomes to underlying factors.⁶⁶

Incomplete specification is not a necessary part of a mandatory social security system. For example, in Chile workers are required to save 10 percent of covered earnings in mutual funds, using the accumulation for an annuity purchase or phased withdrawals after reaching benefit eligibility. Thus the workings of the system are fully specified in terms of economic outcomes.⁶⁷ This does not imply that the Chilean government will never change the rules of the system. Indeed, it has made frequent changes in some details. But it does mean that we can analyze the outcomes of the current system under the assumption of no further legislation without being internally inconsistent. We cannot do that for the United States—there are states of nature that require some legislative change, indeed the probability of such a future need is very high today.

The Chilean approach of a fully funded defined contribution system is not the only way to have a fully specified system. Sweden has one too. In Sweden, the payroll tax rate is 18.5 percent. While 2.5 percent of payroll goes into fully funded individual accounts, 16 percent is used for a partially funded system, called a notional defined contribution system (NDC). An NDC system mimics a fully funded defined contribution system in that it accumulates a notional balance for each worker that increases each year by taxes paid and a notional interest rate.⁶⁸ At retirement, this balance is converted into an annuity based on the

life expectancy of that cohort and the same notional interest rate.⁶⁹ The notional interest rate is set administratively (with automatic adjustment), not by returns realized on assets held. In this way an NDC system mimics a defined benefit system. Thus it is very much a hybrid. Whether this system is referred to as a defined benefit system or an unfunded defined contribution system matters since the vocabulary with which a system is described can influence the politics of both creation and adaptation.

By itself a well-structured NDC system, with a decent size buffer stock of assets, will have little probability of needing legislative intervention as long as economic growth is large enough. Even so, the Swedes have gone further by introducing an automatic balancing system. I will not digress to describe the Swedish automatic balance rules. It suffices to say that if economic growth is sufficiently slow, the notional interest rate is automatically lowered—reducing both benefits in payment and future benefits in response to this slower rate of growth. Thus the Swedish system can be analyzed for an indefinite future without an assumption about the structure of future legislation, so one does not need a fully funded system to have that property. Sweden, like Chile, puts all of the risk of future outcomes on the side of benefits and none on the side of taxes.⁷⁰

Some simple ways for pretty much ensuring automatic balance can illustrate alternative approaches. In the French and German pension systems, workers accumulate “points” based on earnings that have been subject to tax. Think of this as a sum of wage-indexed wages over a worker’s career. The accumulations of points determine relative pensions for retirees. Unlike what is actually done in France and Germany, points could be converted into cash benefits by automatically adjusting the value of a point to exhaust available revenues. Conversely, we could think of adjusting the tax

⁶⁶ For example, if we want to project a legislative response to a possible drop in fertility, we do not have much of a database for evaluating the relationship. With underlying country differences being very significant in social security politics, we may have basically one data point.

⁶⁷ When workers purchase real annuities from insurance companies, there is always the possibility that the insurance companies will become unable to pay the contracted amounts. But even with recognition of this possibility, we still have a fully specified outcome—as we do in models with incomplete markets and bankruptcy rules.

⁶⁸ That is, unlike the United States where benefits depend on earnings subject to payroll tax, in Sweden benefits depend on taxes paid. Since the Swedes currently seem determined not to change the tax rates this difference is likely to have little consequence for the future.

⁶⁹ Also automatically adjusted is the relationship between the level of benefits and the age at which an individual starts them. There is not automatic adjustment for the earliest age at which retirement benefits can be claimed.

⁷⁰ To some, this is the heart of a defined contribution system, rather than a relationship involving realized rates of return on assets actually owned.

rate each year to produce enough revenues to cover expenditures for given values of points.⁷¹ Both types of adjustment need a small buffer stock of assets (or borrowing ability) because of lags between setting benefit or tax rules and the determination of actual expenditures and revenues.⁷²

U.S. Social Security uses price and wage indexing in the determination of both benefits and the payroll tax base.⁷³ This reliance on automatic adjustments decreases the frequency of the need for legislation.⁷⁴ One popular proposal is to extend automatic adjustments to include an adjustment for life expectancy. Such a change would play two roles—one is to have an automatic adjustment rather than legislating in anticipation of or in response to life expectancy changes. The other is to decrease the actuarial imbalance in a way that may be easier politically than comparable direct changes.

But what mix of benefit and revenue changes is the best response to increased life expectancy?⁷⁵

⁷¹ To some, this is the heart of a defined benefit system, including possibly placing the risk outside the labor market, as can be done if the risk is shifted to corporations or general revenues.

⁷² One difference is that if we attempt to increase tax revenues (as opposed to lowering benefit payments), we face the risk of exceeding the maximum that could be collected (i.e., moving to the wrong side of the Laffer curve). Presumably, with a sensible execution of this approach, this risk would be so low as not to be a problem. Adjustment possibilities are more complex than just some combination of tax increases and benefit decreases (or the converses). In the presence of a projected deficit, benefit reduction can be large enough to lower tax rates and tax increases can be large enough to raise benefits. Recognizing more complexity, some tax rates could go up while others go down and benefits for some groups could go up while benefits for other groups go down. Indeed, several proposed reform plans include increased benefits for some vulnerable groups along with general benefit cuts.

⁷³ The current indexing is not complete—there is no adjustment of benefits for inflation between the years a worker is 60 and 62. This gap should be closed.

⁷⁴ Indeed, the 1972 automatic indexation for inflation (which was done incorrectly) was an attempt to codify how Congress had been behaving, thereby reducing the frequency of the need to legislate. The automatic indexing was done incorrectly because congressional actions had been unsatisfactory in structure, without this being as apparent as when the changes became automatic and inflation increased.

⁷⁵ We could also consider an automatic adjustment of the earliest eligibility age and of the actuarial adjustments for the age at which benefits start. The latter, but not the former, is included in Sweden. Indexing the earliest eligibility age is complex since the sizes of the groups

Part of an approach to this question is to ask how individual lifetime plans should vary with life expectancy. This depends on how the potential earnings trajectory and the difficulty (disutility) of work change along with life expectancy. If both opportunities and difficulties in a year depended on the proportional position of that year relative to life expectancy (and mortality rates also depended on relative age), then all of an optimal individual adjustment would come in working longer. That is, optimal work would be a fixed fraction of life expectancy. The change in Social Security with the same pattern has all of the adjustment in lower benefits for each age of retirement.

However, I suspect that the proportional case assumes too large a change in both earnings opportunities and difficulty in work relative to life expectancy. If the optimal outcome for an individual were to work a smaller percentage of life expectancy, then a sensible approach would spread the implied drop in lifetime consumption over both pre- and postretirement years. Decreased preretirement consumption corresponds to an increase in the Social Security tax rate. In historic data, where the steady growth in life expectancy has been accompanied by a steady growth in real earnings, we have a steady decline in the percentage of life expectancy worked. This suggests a mix of tax and benefit changes since we do expect a continued correlation between life expectancies and earnings levels.⁷⁶ I also believe that, at least among academic economists, the life cycle of productivity relates to more than just health and it is unclear how such other factors are correlated with life expectancy. I think an automatic adjustment for life expectancy that included adjustment in both benefits and tax rates would be a good idea.

Should we have more automatic adjustments

helped and hurt by an increase are not likely to be simply related to life expectancy.

⁷⁶ Automatic adjustment of benefits for life expectancy is naturally done on a cohort basis, while any adjustment in taxes is naturally done on a yearly basis. Thus more rapid increases in life expectancy would fall differently on different cohorts when taxes are included in the adjustment than when they are not. My plan with Orszag (2004) takes the approach of a mix of automatic tax and benefit changes for life expectancy, while Model 3 of the Commission to Strengthen Social Security does all its automatic adjustment for life expectancy on the side of benefits.

and so even less pressure for legislation? For example, we could use additional adjustments depending on real wage growth. Or we could go directly to automatic adjustments based on overall financial balance so Congress never again needed to legislate.⁷⁷ Such indexing would need to choose the mix of revenue and benefit changes in the response to imbalance. It strikes me as implausible that a system with a sensible tax rate would want to do all of the adjustment on the side of benefits.⁷⁸ That is, it seems likely that the optimal size of a social security system relative to the economy varies with the same factors that affect actuarial balance.

Relying on fully automatic adjustment rather than assuming there will be periodic new legislation bears some similarity to a familiar distinction from macroeconomics—rules vs. discretion for monetary policy. Parallel issues include the concern about setting rules without fully knowing how the economy adjusts to the policy actions and recognition that the economy may evolve so that currently good rules may become less so in the future. But there are also different issues. Social Security set up for the indefinite future involves a level of detail complexity that seems higher than setting rules for the Fed. Moreover, Congress could invite the Fed to set out a rule it will then follow. Thus we need to ask whether Congress would do a better job in setting out rules once and for all rather than adjusting them from time to time. While legislating from time to time is an inherently easier intellectual problem, we need to be concerned about the asymmetries in the political ease of legislation addressing surpluses and deficits, an asymmetry that is reduced by legislating automatic adjustments. Also there may be more similarity across the

political spectrum in normative evaluations of the impacts of monetary policy than of the evaluations of the sizes of taxes and benefits for different workers and family structures. As with monetary policy, I think that some discretion can improve outcomes.

In considering possible automatic adjustments, one can look at how adjustment is currently debated and how it was done before. In our last major reform in 1983, there was an explicit sense of balancing benefit and revenue changes (Paul C. Light, 1985). Currently, we view both benefit reductions and tax revenue increases as candidates to contribute to restoring a projected position of financial balance. The Commission appointed by President Bush put forth two plans which restored actuarial balance (Commission to Strengthen Social Security, 2002). One of them included new dedicated revenues, and both of them included large transfers of general revenues, which one cannot help but think of as in large part coming from new revenues and not just spending decreases and certainly not benefit decreases. The plan that Orszag and I have put forth explicitly divides some of the proposed changes for restoring actuarial balance (both one-time changes and new automatic changes) between revenue changes and benefit changes.

A. *Fully Funded Defined Contribution and Partially Funded Defined Benefit*

The parallel to the workings of the Arrow-Debreu model and the completeness of the specification makes economists more comfortable thinking in terms of mandated fully funded defined contribution systems than the type of partially funded defined benefit system we have. So, I want to draw out some comparisons. One is that portfolio risk in a mandatory fully funded defined contribution system is highly correlated with the portfolio risk of the rest of individual retirement savings. Thus the increased use of defined contribution private pensions raises the value of a defined benefit Social Security system relative to individual accounts.⁷⁹ In contrast, a

⁷⁷ With this approach, Social Security would become fully specified and so easier to analyze and more in keeping with Arrow-Debreu thinking. But, making it easier to analyze does not necessarily make it better. Just as mathematical convenience, for example from additive preferences, while convenient for theoretical analysis, does not necessarily add to empirical reliability.

⁷⁸ If the tax rate is thought to be too high and politically can not be lowered, then doing the adaptation to anticipated higher cost outcomes fully in terms of lower benefits may make sense as a political fallback. I, for one, do not think the current tax rate and replacement rates in the United States are too high.

⁷⁹ Social Security needs to be considered in the context of all retirement income provision, not just by itself, recognizing the great diversity in the extent to which people have private sources of retirement income. The Social Security reform debate has recognized non-Social Security retirement

system that is less than fully funded will have less correlation with the returns on private retirement savings, as it adjusts benefits in response to the growth of tax revenues as well as the returns on whatever assets are held.⁸⁰ This comparison holds even with initial benefits fully automatically adjusted for the actuarial position—returns on assets and the growth of taxable earnings are only partially correlated. Thus portfolio diversification considerations suggest an advantage to having (at least) some underfunding in Social Security to complement private savings.

This diversification advantage comes with the redistribution to earlier cohorts that is inherent in a less than fully funded system.⁸¹ Thus one can readily argue for a Pareto gain (*ex ante*) from moving from a fully funded system to a partially funded system with the risk associated with the incomplete funding falling on benefits.⁸² Note that the reverse argument does not work—just adding funding to an unfunded system that provides all of retirement benefits will not generate a Pareto improvement. The gain from diversification plays out over time, while the redistribution required to build up funding hurts the oldest cohorts who provide the funding and are affected by the diversification argument very little or not at all.⁸³ Thus the diversification

argument by itself does not lead to the possibility of a Pareto gain from adding funding to an unfunded system, just from reducing funding of a fully funded defined contribution system.

The comparison above assumed all of the response in the unfunded system occurred in benefits. By having some of the response to aggregate shocks fall on taxes, a less than fully funded system is capable of doing additional risk sharing across generations that does not occur with a fully funded defined contribution system (Douglas Gale, 1990). Of course how good a job Congress does in adapting such a system (whether done automatically or by repeated legislation) is a further issue that must be recognized. Thus, the current system provides insurance for individual earnings risk through the progressive benefit formula and the retirement test and provides insurance for aggregate earnings risk through the defined benefit structure with less than full funding.

My broad conclusion here is that the absence of a complete specification of Social Security is not by itself an argument that there is anything wrong with our current approach.

VII. Concluding Remarks

Occasionally, I run into people who believe that no one in his right mind would design a retirement income system like the one we have. Some of the details do seem far from satisfactory to me. However, looking at the big picture, this structure makes sense. Mandated savings makes sense if you think that many workers would not provide themselves a reasonable replacement rate. This is not just an issue of avoiding poverty, but one that extends quite far up the income distribution. Mandating annuitization makes sense if you think that workers do not adequately understand the value of annuities. Protection of spouses and children makes sense if you think that many workers would not do that adequately. Relating benefits to a measure of lifetime earnings surely makes sense. A progressive benefit formula makes sense to provide higher replacement rates for lower earners, in order to supplement annual income taxation

income issues in that some reform proposals have been packaged with increased tax incentives for individual retirement savings. The debate has not included the possibility of a mandatory widening of employer-provided retirement income (with a government-provided default) along the lines of the recent reform in Australia and earlier discussion in the United States (President's Commission on Pension Policy, 1981). The diversity in private savings is also relevant for considering the right size for mandated benefit provision.

⁸⁰ For example, with an NDC system, the notional assets in an account earn a notional rate of return that might be set to equal the growth rate of the wage bill (if other factors are not too strong). Then, one can consider the correlation between the return on assets and the growth rate of the wage bill over different time horizons. Since the internal rate of return in a pay-as-you-go system is related to the timing of tax payments and benefits as well as the growth rate of the wage bill, a more complicated correlation could be examined.

⁸¹ In principle, diversification could also be accomplished by swaps of different tax revenues.

⁸² Throughout this address, I have assumed that the interest rate is above the growth of wages, so that the economy is not on the wrong side of the golden-rule level of capital.

⁸³ This logic is clear in a two-period model. With more periods and externalities, one might find a complex way to

benefit everyone, taking advantage of the externalities to offset the payment for funding. But I have not seen anyone show such a possibility.

as part of lifetime earnings insurance and redistribution, to offset some of the redistributive effects of uniform annuitization, and to address the low antipoverty protection for the elderly (SSI) in contrast with other advanced countries. The retirement test at some ages makes sense.

Having redistributed to earlier cohorts, spreading the implied cost over the indefinite future (not fully funding Social Security) makes sense and incomplete funding contributes to risk sharing across cohorts. Relying on a mix of a smaller mandatory system than is common in Europe and voluntary private supplementation makes sense, even though the voluntary system is so incomplete in its coverage (and in need of improved regulation). This is not to say there are not other approaches that have led to systems that function reasonably well. It is just to say there is no need for radical reform in order to have a good system—just a need to put the program on a stronger financial footing while improving the benefit structure at the same time.

I chose to write about why Social Security is better than many people think rather than why trying a radical reform would be worse than many people think. A preference for a radical reform can reflect different values and different political predictions from mine. However, I think that much of the apparent appeal of radical reform lies in the implausible implicit assumption that such a reform will pass into legislation untouched by political hands, making for a faulty comparison with the current system which has flaws introduced and preserved by the political system. A major reason for my concern about radical reform is the potential for ill-advised design, driven by political ideology rather than a realistic assessment of likely outcomes. A better goal than seeking radical reform is trying to improve the highly satisfactory current structure.

REFERENCES

- Ameriks, John.** “Recent Trends in the Selection of Retirement Income Streams among TIAA-CREF Participants.” *Research Dialogue*, 74, TIAA-CREF Institute, 2002.
- Arnold, R. Douglas.** *The logic of congressional action*. New Haven, CT: Yale University Press, 1990.
- Bernheim, B. Douglas.** “The Economic Effects of Social Security: Towards a Reconciliation of Theory and Measurement.” *Journal of Public Economics*, August 1987, 33(3), pp. 273–304.
- _____. “How Strong Are Bequest Motives? Evidence Based on Estimates of the Demand for Life Insurance and Annuities.” *Journal of Political Economy*, October 1991, 99(5), pp. 899–927.
- Beveridge, W. H.** *Social insurance and allied services*. Cmd 6404. London: Her Majesty’s Stationery Office, 1942.
- Board of Trustees of Social Security and Medicare.** *Status of the Social Security and Medicare programs annual report*. Washington, DC: Social Security Administration, 2003.
- Brown, Jeffrey.** “Redistribution and Insurance: Mandatory Annuitization with Mortality Heterogeneity.” *Journal of Risk and Insurance*, February 2003, 70(1), pp. 19–43.
- Coile, Courtney and Gruber, Jonathan.** “Social Security Incentives for Retirement,” in David Wise, ed., *Themes in the economics of aging*. Chicago: University of Chicago Press, 2001, pp. 311–41.
- Commission to Strengthen Social Security.** Final Report, 2002; www.csss.gov/reports/Final_report.pdf.
- Davidoff, Thomas; Brown, Jeffrey and Diamond, Peter.** “Annuities and Welfare.” Working paper, Massachusetts Institute of Technology, 2003.
- Diamond, Peter A., ed.** *Issues in privatizing Social Security: Report of an expert panel of the National Academy of Social Insurance*. Cambridge, MA: MIT Press, 1999.
- _____. *Taxation, incomplete markets and social security: The 2000 Munich lectures*. Cambridge, MA: MIT Press, 2002.
- Diamond, Peter A. and Orszag, Peter R.** *Saving Social Security: A balanced approach*. Washington, DC: Brookings Institution Press, 2004.
- Gale, Douglas.** “The Efficient Design of Public Debt,” in Rudiger Dornbusch and Mario Draghi, eds., *Public debt management: Theory and history*. Cambridge: Cambridge University Press, 1990, pp. 14–47.
- Holden, Karen and Zick, Cathleen.** “Insuring Against the Consequences of Widowhood in a Reformed Social Security System,” in R. Douglas Arnold, Michael J. Graetz, and Alicia H. Munnell, eds., *Framing the Social*

- Security debate: Values, politics and economics*. Washington, DC: National Academy of Social Insurance, distributed by Brookings Institution Press, 1998, pp. 157–70.
- Leimer, Dean R.** “Cohort-Specific Measures of Lifetime Net Social Security Transfers.” Social Security Administration Office of Research and Statistics Working Paper No. 59, February 1994.
- Light, Paul C.** *Artful work: The politics of Social Security reform*. New York: Random House, 1985.
- Lindbeck, Assar and Persson, Mats.** “The Gains from Pension Reform.” *Journal of Economic Literature*, March 2003, 41(1), pp. 74–112.
- Marshall, Alfred.** *Principles of economics*, 8th Ed. New York: Macmillan Company, 1948.
- Merton, Robert C.** “On the Role of Social Security as a Means for Efficient Risk Sharing in an Economy Where Human Capital Is Not Tradable,” in Zvi Bodie and John Shoven, eds., *Financial aspects of the US pension system*. Chicago: University of Chicago Press, 1983, pp. 325–58.
- Munnell, Alicia and Sundén, Annika.** *Coming up short: The challenge of 401(k) plans*. Washington, DC: Brookings Institution Press, 2004.
- President’s Commission on Pension Policy.** *Coming of age: Toward a national retirement income policy*. 1981.
- Shiller, Robert J.** *Macro markets: Creating institutions for managing society’s largest economic risks*. Oxford: Clarendon Press, 1993.
- Social Security Administration**, Office of Policy, Office of Research, Evaluation and Statistics. “Fast Facts and Figures about Social Security.” Washington, DC: Social Security Administration, 2003.
- Yaari, Menahem.** “Uncertain Lifetime, Life Insurance, and the Theory of the Consumer.” *Review of Economic Studies*, April 1965, 32(2), pp. 137–50.