Failures of Contingent Reasoning in Annuitization Decisions

Erzo F.P. Luttmer ⓡ Priscila de Oliveira ⓡ Dmitry Taubinsky*

March 28, 2025

Abstract

This paper studies psychological biases in take-up of annuities, using an incentivized experiment with a probability-based sample (N = 3,038). Choosing an annuity was payoff-maximizing in the experiment at all prices, but take-up was incomplete and price elastic. Reformulating decisions as insurance against a "bad" outcome rather than insurance against "longevity risk" did not increase take-up. Instead, we find substantial failures of contingent reasoning: participants under-appreciated how annuitization mitigated the need for less-efficient means of saving for retirement. Increasing the salience of the interaction with savings decisions, or eliminating the need to think through this interaction altogether, substantially increased annuity take-up.

JEL Codes: D14, D15, D9, G5, J26

Keywords: Annuity puzzle, retirement savings, aging, behavioral public economics, contingent reasoning

---

*Luttmer: Dartmouth College and NBER. Erzo.FP.Luttmer@dartmouth.edu. de Oliveira: Nova School of Business and Economics. priscila.oliveira@novasbe.pt. Taubinsky: UC Berkeley and NBER. dmitry.taubinsky@berkeley.edu. We thank Jeffrey Brown, Olivia Mitchell, and Stephen Zeldes, and participants at seminars and conferences for helpful comments. We are grateful to the Alfred P. Sloan Foundation for funding (Grant#: G-2018-11101). We thank the AmeriSpeak team, and in particular Dan Constanzo and LeeAnn McCabe for help with implementing the experiment. We are grateful to Carl Harris, Isabel Galea, Chenxi Jiang, Yikun (Allison) Jiang, Amina Ospan, Umair Shahbaz, Afras Sial, and Laila Voss for excellent research assistance. Our experiment was determined exempt by the Berkeley IRB (#2019-12-12779) and by the Dartmouth IRB (#STUDY00031949). The analysis plan was pre-registered on the AEA RCT registry, RCT ID AEARCTR-0006903.

Optimally preparing for retirement requires people to navigate a complex set of decisions with intricate dynamics, nuanced interactions between different financial instruments, and many sources of uncertainty. One important source of uncertainty is longevity—the longer one lives, the more resources one needs for old age. As articulated by Yaari (1965), annuities provide insurance against this "longevity risk."

Consider, for example, an individual who in period 1 is "young" and earns income, in period 2 is "old" and does not earn income, and survives to period 2 with 50% chance. Normalizing the interest rate to 0, suppose that it is optimal to save amount s of period-1 earnings for retirement if no annuities are available. Then, an actuarially fair annuity that costs $p = s/2$ and pays out $a = s$ in period 2 if the individual survives increases consumption. By purchasing the annuity, the individual maintains the same level of retirement consumption in the event of survival, but does so at only half the cost to period-1 consumption. In fact, the annuity need not be actuarially fair: purchasing any annuity at price $p < s$ and payout $a \geq s$ increases consumption in both periods.

Despite the strong theoretical arguments for at least partial annuitization under broad conditions (Yaari, 1965; Davidoff et al., 2005), the demand for annuities that is observed in reality is remarkably low (Mitchell et al., 2011; Poterba et al., 2011). For example, Arapakis and Wettstein (2023) report that only 12% of retirees with $\$100,000$ or more to invest own an annuity. Brown (2009) reviews a host of "rational" explanations for the low take-up—including adverse selection, bequest motives, and uninsurable late-in-life expenditure shocks (e.g., Ameriks et al., 2011; Lockwood, 2012; Reichling and Smetters, 2015; Peijnenburg et al., 2017)—as well as psychological biases that may inhibit take-up. Possible psychological frictions include lack of financial literacy leading to an aversion to buying complex products, choice bracketing causing annuities to be seen as a gamble, and a heuristic that insurance is only for "bad outcomes." However, observational data is inconclusive about the role of standard economic preferences versus psychological frictions, as well as which of these is most relevant.

This paper analyzes a tightly controlled, incentivized experiment that uses variations of the example above to study how and why people may fail to optimally allocate resources for retirement. We use a simple experimental setting where, in contrast to prior work, we can unambiguously classify not choosing an annuity as a mistake. In our setting, choosing an annuity was always payoff maximizing, and in fact first-order stochastically dominant for all but the most suboptimal savings choices. The stochastic dominance feature of our design implies that not choosing an annuity was an unambiguous mistake for any assumption about the nature of risk preferences over modest monetary amounts.

Participants played a game in which they allocated tokens between stage 1 and stage 2,

where stage 2 was reached with 50% chance. Tokens in each stage were converted to rewards by a concave function: participants had to maintain at least 40 tokens in each stage, received \$0.25 per token for each additional token from 41 to 80, and received no additional reward for additional tokens beyond that. In the Benchmark condition, participants first chose whether or not to purchase an annuity for stage 2 at a price that was less than actuarially fair, and then decided on the level of savings. The variation in the other conditions was designed to evaluate two key psychological biases that might depress annuity take-up. We ensured that participants understood the game by providing clear explanations that included examples and comprehension checks, and by allowing participants to continue only if they passed most of the comprehension checks.

Our first hypothesized bias—first described in Brown (2009)—was that a combination of narrow bracketing and other heuristics employed in choice under risk (Tversky and Kahneman, 1981; Read et al., 1999; Rabin and Weizsäcker, 2009) make it counterintuitive to pay for a financial instrument that not only has uncertain returns but also pays out in a "good" state of the world, such as longer life. In most cases where people purchase insurance, the state with higher marginal utility from money is a "bad" state with lower absolute utility. To test this hypothesis, we designed a condition in which the payoffs to people's decisions were identical, but the stage-2 correlation between the marginal value of a token and the absolute payout was reversed. In this Reverse-Correlation condition, participants always reached stage 2, but the uncertainty was whether they would receive additional income of over 80 tokens (50% chance) or not receive any additional income (50% chance). The annuity was described as insurance against losing their stage-2 income. Because additional tokens above 80 generated no reward, the impact on payoffs of any combination of a savings and annuity choice was identical to the Benchmark condition.

Our second hypothesized bias was *failures of contingent reasoning*. A growing body of work in psychology and economics shows that people struggle with the kind of hypothetical thinking necessary for working through decision trees in dynamic and uncertain environments (e.g., Shafir and Tversky, 1992; Esponda and Vespa, 2014; Li, 2017; Esponda and Vespa, 2024; Martínez-Marquina et al., 2019). Most starkly, this work shows that difficulties with contingent reasoning can lead to violations of dominance: a person who prefers alternative a over alternative b in *every* state of the world may nevertheless choose b over a when the state of the world is not revealed. Thus, even if the annuity is utility-maximizing, as in our experiment, difficulties with contingent reasoning could lead people to under-appreciate its value.[1]

[1]This difficulty with thinking through contingencies is distinct from complexity aversion. On the one hand, people might be averse to deterministic payoffs described by complex formulas, which would generate complexity aversion but not failures of contingent reasoning. On the other hand, failure of contingent

We designed four different Salient-Contingencies conditions that made the consequences of choosing an annuity easier to grasp. In the first condition, people made their decisions about savings—both for the case in which they have an annuity and for the case in which they don't—before deciding whether to get the annuity or not. The second condition built on the first by clarifying the levels of savings (previously chosen by the participants) tied to their annuity decision, and the resulting number of tokens in each stage. The third condition removed all context appearing in the second condition and simply offered people a choice over tokens in stage 1 and stage 2. The fourth condition modified the third condition by adjusting savings associated with the annuity so that tokens in stage 2 were the same with and without the annuity. The adjustment made the annuity strictly stochastically dominant for the 22% of participants for whom the annuity was not strictly stochastically dominant at their original savings choice.

Not having an annuity was the status quo in the Benchmark condition, and participants needed to decide whether or not to buy the annuity. However, the Salient-Contingencies conditions lacked a status quo because this enabled us to better spell out the contingencies for each choice. To assess status-quo effects, we included a condition in which participants made a direct choice between buying and not buying the annuity, but where the contingencies were not made salient. Additionally, to provide a point of comparison for the effect sizes of our other variations, we introduced a condition where the price of the annuity was lowered to be better than actuarily fair.

We present four main sets of results. First, we find that even though the annuity was strictly stochastically dominant for 78% of participants and payoff-maximizing for all, only 71% took it up in the Benchmark condition. However, take-up increased to 88% when the annuity price was lowered to be better than actuarily fair. That participants were elastic to the price decrease indicates that they were not just heuristically avoiding the annuity altogether, but instead misconstruing its value.

Second, and contrary to our initial hypothesis, take-up was *lower* in the Reverse-Correlation condition. This suggests that the notion of longevity insurance is not unnatural to participants per se, and in fact may be more natural than other forms of insurance.

Third, our Salient-Contingencies conditions increased take-up from 71% to 83%, on average. The majority of this effect was not due to removing the non-annuity status quo, as this manipulation increased take-up to only 75%. Simply putting the savings decisions before the annuity decisions increased take-up to 81%, while collapsing the decision tree by fully

reasoning can occur even if annuities look simpler than alternative savings vehicles under consideration, if those other savings vehicles involve fewer contingencies that affect payouts in a way that is strongly correlated with the owner's marginal utility of income.

spelling out the consequences of choosing an annuity increased take-up to 87%. Ensuring that annuities led to stochastically-dominant payoffs had no additional effect, suggesting that a combination of suboptimal choice of additional savings and extreme levels of risk aversion was not contributing to incomplete take-up.

Fourth, we find that our Salient-Contingencies treatments had the *largest* effect on people with the highest levels of financial literacy and comprehension of our experimental setting. This suggests that failures of contingent reasoning in the annuity context are a deep-seated bias that does not just affect people who are the least financially literate or the least motivated to optimize their choices. Together with our protocol to allow participants to complete the experiment only if they passed the majority of the comprehension checks, this also is a clear indication that simple confusion about the experimental setting did not drive any of our main conclusions.

Generating a setting in which choosing annuities is unambiguously optimal requires highly controlled experimental designs that are difficult to implement in more naturalistic settings. We designed our experiment to provide the first clean test of systematic biases in annuity take-up, but a natural limitation of this approach is that our specific quantitative results are hard to extrapolate to real-world settings. Instead, the goal of our experiment is to provide proof-of-concept that systematic mistakes can depress annuity take-up, and to provide qualitative insights into the nature of such mistakes. In our concluding section, we further discuss external validity and how future work can build on our results to explore biases in annuity take-up in more naturalistic settings.

To our knowledge, our paper is the first to develop a design where not purchasing an annuity can be classified as an unambiguous mistake. It is also the first paper to empirically study how "reverse correlation" and failures of contingent reasoning can affect annuity take-up. The only papers conducting controlled, incentivized experiments on annuity choice (Agnew et al., 2008; Gazzale and Walker, 2009) focused on the role of status-quo bias and demographic covariates, such as gender. More papers have used surveys to investigate possible behavioral biases in annuity choice. Brown et al. (2008), Brown et al. (2013), Beshears et al. (2014), and Brown et al. (2016) show that manipulating the language and framing of an annuity decision can alter people's stated preferences for take-up, which is suggestive of behavioral biases. However, as in observational data, the normative benchmark for take-up is ambiguous in these studies, and thus it is unclear whether people *under*-annuitize due to behavioral biases or whether framing manipulations can lead people to *over*-annuitize. Brown et al. (2017) and Brown et al. (2021) provide evidence of a buy-sell spread in hypothetical annuity transactions, and show that it is mediated by the complexity of a decision and participants' financial sophistication—but do not provide definitive evidence of whether

this leads people to *under*-annuitize. These two papers are consistent with our secondary finding of moderate status-quo bias.

More broadly, our work contributes to literatures on bounded rationality in public economics and household finance (see Bernheim and Taubinsky, 2018 and Beshears et al., 2018, respectively, for reviews). The novel finding of a failure of contingent reasoning in our simple experimental rendition of a retirement savings decision suggests that it would be valuable to investigate whether this bias matters in other settings involving dynamic consumption decisions with multiple financial instruments (Chakraborty and Kendall, 2022).

1 Experimental Design

1.1 Annuity and Savings Decisions

Basic structure. Our "Life-Planning Game" was based around the annuity choice model of Davidoff et al. (2005). This model has no bequest motives, which could reduce the demand for annuities. The game described the two periods as "stage 1 - when you're young" and "stage 2 - when you're old." Stage 2 had two equally likely outcomes: "you survive" or "you don't survive." Participants received an endowment of "90 tokens of income" in stage 1, some of which could be used to "buy an annuity" and some of which needed to be "saved" for stage 2. If the stage-2 outcome was "you survive," participants got the tokens they had saved for stage 2 and received tokens from the annuity (if they had one). If the stage-2 outcome was "you don't survive," participants did not get tokens from their annuity in stage 2 (if they had one) and could not use the tokens from their savings in stage 2.

There were two ways of transferring tokens to stage 2: buying an annuity and saving tokens. Therefore, participants made two types of decisions in the experiment. In *annuity decisions,* participants chose whether or not to buy an annuity at two possible prices. The annuity cost either 10 or 20 tokens in stage 1 (low and high price, respectively) and always payed out 30 tokens in stage 2 when alive and 0 tokens otherwise. Therefore, each token transferred using the annuity generated a stage-2 expected value of 1.5 or 0.75 tokens, respectively. In *savings decisions,* we showed participants how many tokens they had in each stage, and asked "How many tokens would you like to save from stage 1 for stage 2?". We asked participants to make a separate savings decision for three different scenarios: (i) not having annuity, in which case they had 90 tokens in stage 1 and 0 tokens in stage 2 when alive, (ii) having a low-price annuity, in which case they had 80 tokens in stage 1 (= 90 minus the cost of 10 for a low-price annuity) and 30 tokens from the annuity in stage 2 when alive, and (iii) having a high-price annuity, in which case they had 70 tokens in stage

1 (= 90 minus the cost of 20 for a high-price annuity) and 30 tokens from the annuity in stage 2 when alive. We elicited savings decisions for these three cases because some of our experimental manipulations, as described below, consisted of asking savings decisions before the participant had made their annuity decision. Hence, we needed to know what savings would be for each possible realization of annuity ownership.

In stage 2, participants received their saved tokens from stage 1 if they "survived." However, if they "didn't survive," they got no tokens, which meant to model the fact that in the absence of bequest motives, people cannot use their savings after death. This implies that each token saved generated a stage-2 expected value of 0.5 tokens, which is less than the expected value of 0.75 or 1.5 tokens generated per token spent on an annuity. Hence, our experiment reflects the observation by Davidoff et al. (2005) that an annuity is a "dominant asset" if wealth after death has no value.

The payout in each decision task was based on the final token allocation in each stage. The first 40 tokens in stage 1 and in the alive state in stage 2 were mandatory and did not generate pay. Tokens 41 to 80 paid \$0.25 each, and tokens above 80 paid \$0 each. If participants tried to save an amount that would result in fewer than 40 tokens in stage 1 or the alive state in stage 2, they were reminded of the 40-token minimum and required to adjust their savings. Analogous to concave utility creating incentives to smooth consumption, this concave payoff structure with a subsistence minimum created the need to transfer tokens from stage 1 to stage 2.[2]

At the end of the study, participants saw which decision task and which outcome for stage 2 ("survive" or "not survive") were randomly selected for payout, what they chose in the selected decision task, and how their resulting bonus payment was calculated.

Optimality of annuity take-up. A participant not purchasing the annuity had to save at least 40 tokens out of their 90-token endowment to obtain the required 40-token minimum in stage 2 if they survived. By saving 40 tokens, they retained 50 tokens for a payoff of \$2.50 from stage 1 (tokens 41 to 50 each payed \$0.25) and a payoff of \$0 from stage 2 (because they had no tokens above 40 if they survived). Saving more than 40 tokens reduced the expected payoff. Not purchasing the annuity therefore resulted in a sure payoff of \$2.50 when savings were chosen optimally. A participant purchasing the annuity needed to save

[2]In principle, we could have made the concave payout functions smooth. We opted for a very simple and transparent piece-wise linear structure for two reasons. First, because this is a design where preferences for consumption smoothing across periods are induced, it was important that the induced smoothing motives were very clear. This is important because outside the lab, the basic notion of consumption smoothing is arguably intuitive, and thus we didn't want participants in our lab experiment to have to solve an algebra problem to appreciate this motive in our setting. Second, we needed payoff functions that have zero marginal utility in some regions to cleanly implement the Reverse Correlation condition that we describe below.

only 10 tokens for stage 2 to reach the required 40 tokens, because the annuity gave 30 tokens upon survival. With the high-price annuity costing 20 tokens, a participant would thus optimally retain $90-20-10 = 60$ tokens in stage 1, for a payoff of \$5.00 (tokens 41 to 60 each pay \$0.25) from stage 1 and no payoff in stage 2. Saving in excess of 10 tokens decreased the expected payoff, and thus purchasing the high-price annuity increased the payoff from \$2.50 to \$5.00 if savings were chosen optimally. While suboptimal savings could increase or reduce the payoff of choosing an annuity, the structure of the life-planning game ensured that choosing the annuity would weakly increase expected payoffs for any saving choices. At the savings levels chosen by participants, choosing a high-price annuity increased expected payoffs by about \$1.60, on average (see Appendix Table A1).

Moreover, purchasing the high-priced annuity was beneficial even with suboptimal savings. Suppose that a participant optimally saved 40 tokens in the absence of the annuity, and sub-optimally saved $s > 10$ tokens with the annuity, thus retaining $90 - 20 - s$ tokens in stage 1, and obtaining $30 + s$ tokens in stage 2 when alive. When $s \leq 20$, the participant obtained a payout of $0.25 \cdot (30 + s - 40) = 0.25 \cdot (s - 10)$ from stage 2 with 50% chance, and obtained a payout of $0.25 \cdot (90 - 20 - s - 40) \geq 2.50$ from stage 1. This generated a lottery that stochastically dominated the sure payoff of \$2.50 obtained in the absence of the annuity.[3] When $s > 20$, the participant's payout was still guaranteed to be at least \$2.50 *in expectation*,[4] though the annuity was not guaranteed to stochastically dominate, as there was a 50% chance of the participant not surviving in stage 2, and only obtaining their stage 1 payout of $0.25 \cdot (30 - s) < 2.50$.

Because the high-price annuity always maximized expected payouts in the experiment, the low-price annuity did as well. Moreover, the low-price annuity generated stochastically-dominant payouts for an even larger set of savings levels. As with the high-price annuity, the participant needed to save only an additional 10 tokens for stage 2 with the low-price annuity. However, with the low-price annuity, as long as the participant saved no more than 30 tokens for stage 2, they obtained a payout of at least \$2.50 from stage 1 alone, thus obtaining a payout distribution that stochastically dominated their payouts in the absence of the annuity (for all possible savings choices in the absence of the annuity). Stochastic dominance could fail only if the participants saved more than an additional 30 tokens for stage 2 in the presence of the annuity.

[3]Note that if the participant saved sub-optimally in the absence of an annuity, their payout in the absence of the annuity would be even lower, and the dominance of the annuity would be strengthened.

[4]To see this, note that $0.25 \cdot (30 - s) + (0.5) \cdot (0.25)(s - 10) = 6.25 - 0.125s$, which is weakly greater than 2.50 when $s \leq 30$. Note that saving more than 30 tokens in the presence of the annuity was not possible, as participants had to leave at least 40 tokens for stage 1.

1.2 Experimental Conditions

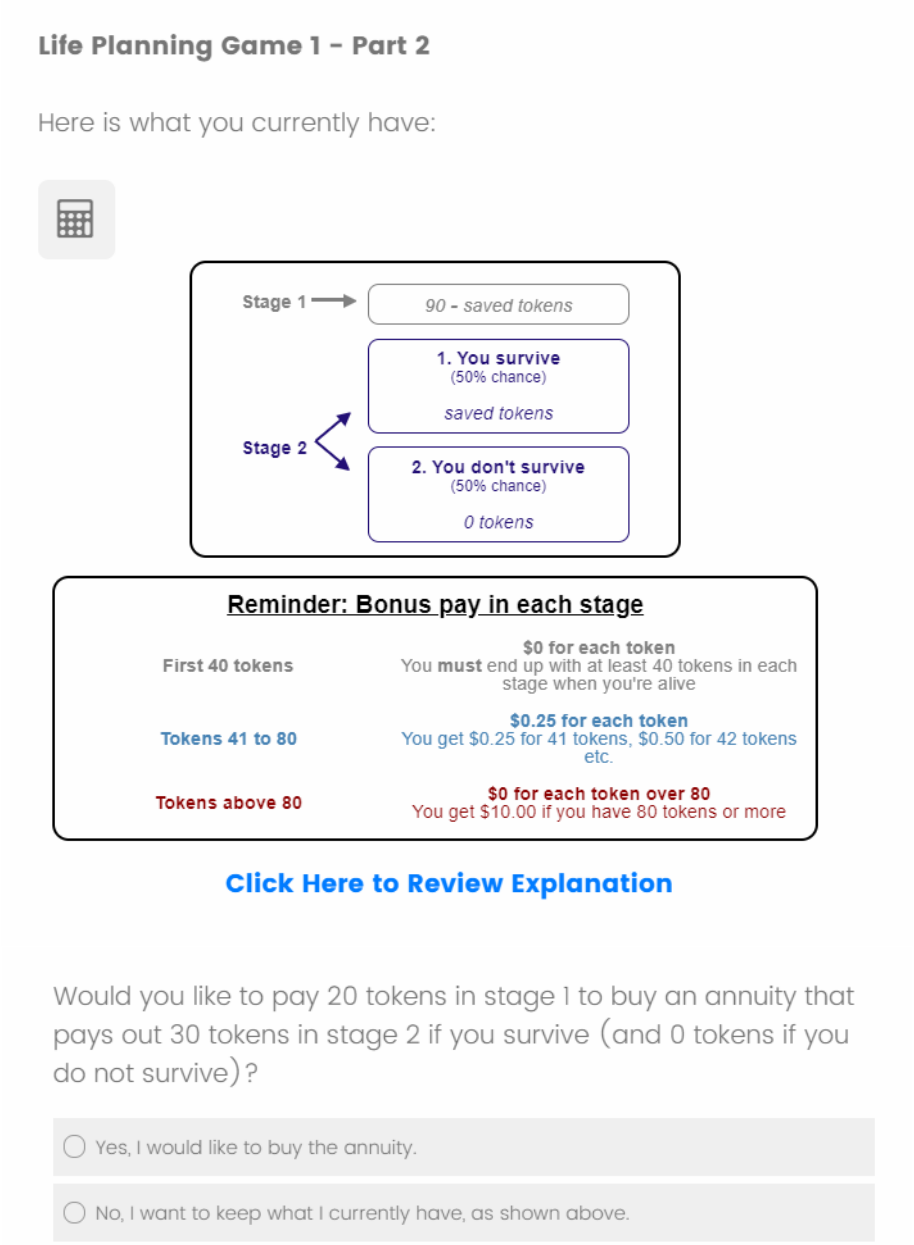
High-price Benchmark condition. The experimental conditions were relative to the "Life-Planning Game" described above using an annuity that costs 20 tokens. This High-price Benchmark case was constructed to resemble the conditions of annuity decisions that people typically face. First, it involved a worse-than-actuarially fair (high-price) annuity, as is the case with most annuities available on the market. Second, participants made the annuity decision before being asked to make the savings decisions for the three scenarios because people typically make annuity decisions without being explicitly asked how much they would save with or without the annuity. Third, as is typically the case with annuity decisions, the status quo was not owning an annuity. Specifically, the annuity choice was presented on two screens. The first screen showed a description and a diagram of what participants "currently have," which displayed an income endowment without an annuity. The second screen read "Here is what you currently have," followed by the same diagram, and asked if they "would like to buy" an annuity at a price of 20 tokens, as in Figure 1.

Low-price Benchmark condition. In the Low-price Benchmark condition, the annuity costs 10 tokens, which is less than the actuarially fair price of 15 tokens. We included this condition to measure responsiveness to price, allowing us evaluate the size of the response to our other experimental conditions using a price metric.

No-Status-Quo condition. This condition was identical to the Benchmark condition, except that it presented an option with an annuity and an option without an annuity next to each other on a single screen without making one the status quo. The options were labeled "Option A" and "Option B," with the position of the annuity option randomized. We created this condition as the comparison for experimental conditions, described below, that had no status quo because there was no natural way of having a status quo in them.

Four Salient-Contingencies conditions. We created these conditions to investigate potential failures of contingent reasoning in annuity decisions. In Salient Contingencies I, the annuity question was presented exactly as in the No-Status-Quo condition, except that we asked it after participants had made their savings choices for each possible realization of annuity ownership (owning no annuity, a low-price annuity, or a high-price annuity). This treatment increased the salience of the savings choices and thereby encouraged respondents to think through the dynamic decision using backwards induction.

Figure 1: Screenshot of decision screen



Notes: This figure presents a screenshot of a decision screen that participants in the High-price Benchmark condition faced, with the "annuity" wording. All experimental instructions are contained in the Supplementary Study Instructions Appendix, https://users.nber.org/~luttmer/StudyInstructionsAppendix.pdf.

Condition Salient Contingencies II (I + Savings Specified) was identical to Salient Contingencies I, except that we additionally told participants how much they had to save if they chose the annuity and how much they had to save if they didn't. These savings levels corresponded to the savings choices they had just made, but we did not point that out. This treatment used the same context-rich setting as (I), but showed in the diagram the final

number of tokens in each stage for each option. This condition eliminated the need to use backwards induction because the diagram showed all consequences of choosing the annuity.

Condition Salient Contingencies III (II + No Context) was a modification of Salient Contingencies II. The diagram was identical, but the text used no contextual terms such as "income" or "annuity," and presented options solely in terms of final tokens received in each stage, without mentioning their sources. We hypothesized that eliminating context could influence behavior because the context could still contribute to the informational complexity of the problem. Because both alternatives still involved multiple stages and uncertainty in stage 2, participants still had to carefully study the options to make the optimal decision. The additional context describing differences in income, savings, and availability of the annuity could have created additional cognitive burden that made it more difficult for participants to focus on the payoff-relevant features of the problem. In this sense, we see Salient Contingencies III as implementing conditions that come as close as practically possible to facilitating participants expressing their *direct judgements* (see, e.g., Bernheim and Taubinsky, 2018; Ambuehl et al., 2022, for a discussion of this concept and approach).

Finally, condition Salient Contingencies IV (III + Dominance) was a modification of Salient Contingencies III designed to ensure that the annuity option stochastically dominates the no-annuity one. We adjusted savings in the option with the annuity such that the number of tokens in stage 2 would be identical to the number of tokens chosen by the participant in the no-annuity option. This adjustment ensured that the annuity option had strictly more tokens in stage 1 because the annuity was a cheaper way of transferring tokens from stage 1 to stage 2 relative to saving. Hence, in this condition, the annuity strictly dominated the no-annuity option, while maintaining the same variance. We included this condition to examine if any lack of take-up in Salient Contingencies III could be due to small-stakes risk aversion.

Reverse-Correlation condition. We designed this condition to study a potential heuristic aversion to allocating income to states of the world in which marginal utility and absolute utility are both high. In this condition, instead of an "annuity," people could buy "insurance" against a loss of stage-2 income. The two possible outcomes in stage 2 were presented as "you don't get income" and "you get income," which were designed to parallel "you survive" and "you don't survive," respectively. In the Benchmark condition, participants got tokens both from the annuity and savings if they "survived" in stage 2. Similarly, in the Reverse-Correlation condition, participants get tokens both from "insurance" and savings if they "didn't get income" in stage 2. In the Benchmark condition, participants got no tokens from the annuity and had no use for their savings if they "didn't survive" in stage 2.

Similarly, in the Reverse-Correlation condition, participants got no tokens from "insurance" and received no value from their savings if they "got income" in stage 2. The reason they received no value from their savings when they got income is that they would have "more than 80 tokens" already in stage 2, so that the marginal utility of an additional token was zero. Thus, feasible savings and the marginal utility of tokens in the Reverse-Correlation treatment were identical to the Benchmark condition. The only difference in the Reverse-Correlation condition is that the marginal utility of tokens was high when the outcome in stage 2 was "bad" (not getting income) rather than "good" (surviving).

1.3 Procedures: Randomization, Incentive-Compatibility, and Comprehension

Wording. For participants not in the Reverse-Correlation conditions, we randomized with equal probability whether the annuity was described as "annuity," "Social Security," or "insurance." The Reverse-Correlation conditions used only the "insurance" wording because that was the most natural word to use. Wording effects are not of primary interest, but because the "insurance" wording was most natural in the Reverse-Correlation conditions, we needed to (i) be able to compare to comparable wording in the other experimental conditions and (ii) therefore make sure that the "insurance" wording did not lead to meaningfully different behavior in these other experimental conditions compared to the "annuity" or "Social Security" wording.

Question blocks. Decision tasks were divided into three blocks: a block with two annuity decisions, a block with one annuity decision, and a block with the three savings decisions. All participants received each of the three blocks, but the exact content and order of the three blocks was randomized across participants. Randomizations were independent of each other unless otherwise noted.

Content of the blocks. The block with two annuity decisions was randomized to one of four conditions with equal probability: Benchmark, No Status Quo, Salient Contingencies I, or Reverse Correlation. Within this block, the annuity question was asked twice, once for the high-price annuity and once for the low-price annuity, with the order of the two questions randomized.

For all participants other than those in the Reverse-Correlation condition, the one-question block contained an annuity decision that was randomized with equal probability to Salient Contingencies II, III, or IV. The annuity was always sold at a high price in this

block.[5]

In the Reverse-Correlation condition, the one-question annuity block had the Reverse Correlation structure (where people either received additional income or not), but this question was always presented in the Salient Contingencies IV condition, and always had a low price. We made this design choice because studying all possible interaction effects of the Reverse-Correlation condition with other conditions was not of primary interest; instead, we wanted to use this question to study ceiling effects by presenting the annuity question in the combination of conditions that we hypothesized would have the largest effect on take-up.

The block with savings decisions asked for savings under three different scenarios: once conditional on not owning an annuity, once conditional on owning a low-price annuity, and once conditional on owning a high-price annuity. The order of the three savings questions was randomized with equal probability.

Order of the blocks and randomization. Table 1 shows how the order of the three blocks depended on the randomization of the content of the two-question annuity block. If the two-question block was randomized to the Benchmark condition (Panel A) or the No-Status-Quo condition (Panel B), the two-question annuity block was asked first, followed by the block with three savings questions, followed by the one-question annuity block in Salient Contingencies II, III, or IV condition. If the two-question block was randomized to the Salient-Contingencies I condition (Panel C), the savings questions was asked first because, by definition, the Salient Contingencies I treatment was to ask the savings questions first. Finally, if the two-question annuity block was randomized to the Reverse-Correlation condition (Panel D), this block was asked first, followed by the savings questions, followed by the one-question annuity block.

The structure of the randomization ensured that each of the nine main experimental conditions specified in the pre-analysis plan would be answered by a quarter of participants in expectation: High-price Benchmark, Low-price Benchmark, High-price No Status Quo, High-price Salient Contingencies I, High-price Salient-Contingencies II, High-price Salient-Contingencies III, High-price Salient-Contingencies IV, High-price Reverse Correlation, and Low-price Reverse-Correlation Salient Contingencies IV.

[5]After collecting data on 1,049 of the 3,038 participants, we added an additional block with a fourth annuity decision to better study price and ceiling effects. This fourth annuity decision was the same as the annuity decision asked in the one-question block, except that it differed in terms of the price of the annuity. To ensure comparability across the entire sample, the additional decision was always the last choice that respondents made. This decision was not specified in our analysis plan and not used in our main analyses. Because it is not used in the main analyses, we do not include it in the description of the experiment below, but Appendix Table A2 shows the randomizations for this additional block.

Table 1: Randomization procedure

Expected Fraction	Nr. of Participants	Savings First	Annuity Block 1		Savings Second	Annuity Block 2	
Panel A: Benchmark							
1/8	426		Low-Price Benchmark	High-Price Benchmark	X	High-Price Salient Cont. II or III or IV	
1/8	396		High-Price Benchmark	Low-Price Benchmark	X	High-Price Salient Cont. II or III or IV	
Panel B: No Status Quo							
1/8	380		Low-Price No Status Quo	High-Price No Status Quo	X	High-Price Salient Cont. II or III or IV	
1/8	370		High-Price No Status Quo	Low-Price No Status Quo	X	High-Price Salient Cont. II or III or IV	
Panel C: Salient Contingencies							
1/16	209	X	Low-Price Salient Contingencies I	High-Price Salient Contingencies I		High-Price Salient Cont. II or III or IV	
1/16	193	X	High-Price Salient Contingencies I	Low-Price Salient Contingencies I		High-Price Salient Cont. II or III or IV	
1/16	195	X	High-Price Salient Cont. II or III or IV			Low-Price Salient Contingencies I	High-Price Salient Contingencies I
1/16	183	X	High-Price Salient Cont. II or III or IV			High-Price Salient Contingencies I	Low-Price Salient Contingencies I
Panel D: Reverse Correlation							
1/8	325		Low-Price Reverse Correlation	High-Price Reverse Correlation	X	Low-Price Reverse-Correlation Salient Contingencies IV	
1/8	361		High-Price Reverse Correlation	Low-Price Reverse Correlation	X	Low-Price Reverse-Correlation Salient Contingencies IV	

Notes: This table describes all of the different experimental cells for annuity take-up decisions used in this paper. Additionally, the three possible wordings ("annuity," "Social Security," "insurance") were randomized at the participant level in the regular arm; the reverse-correlation arm used only the "insurance" wording.

As Table 1 illustrates, comparisons across High-price Salient Contingencies II, III and IV are between-participant, since any participant only faced at most one of these three conditions. Comparisons of High-price Salient Contingencies II, III and IV with High-price Benchmark, High-price No Status Quo or High-price Salient Contingencies I involve both

within- and between-participant variation. Comparisons across High-price Benchmark, High-price No Status Quo, High-price Salient Contingencies I and High-price Reverse Correlation are between participants. All price variation analyzed below is for the Benchmark condition and is within participants.

While in principle we could have had more of the experimental conditions randomized within-participant, we did not do so to keep the experiment as short and simple as possible. Having participants complete additional and more varied annuity decisions would increase the length of the experiment, and could more generally make the experience more cognitively taxing.

Appendix Table A2 contains additional details on randomization, the order of the blocks, and cell sizes.

Incentive compatibility. All decisions in the experiment were incentive compatible. Before making any decision, participants were informed that at the end of the experiment one of the decisions would be randomly selected for payout and that their bonus pay would be determined by their choice in that particular decision. The bonus averaged $5.17, and was paid in addition to the base pay of $2.00.

Since any choice could be selected for payout, participants were always incentivized to select the utility-maximizing option. If a savings decision was selected for payout, the participant's final token allocation in each stage of that decision determined their bonus pay, according to the token-to-dollar conversion. If an annuity decision was selected for payout, the participant's savings choice corresponding to that particular annuity decision was used to determine the final token allocation across stages and, consequently, the bonus pay.

Comprehension questions. Before participants faced any decisions in our experiment, we provided them with thorough explanations of the "Life Planning Game," including graphical examples, step-by-step explanations of how to read the diagrams, and numerical examples of how to calculate final payoffs. Importantly, in all decisions tasks, participants could always click on a link to see a concise version of the example and explanation again.

To encourage participants to internalize the detailed explanations, participants faced seven comprehension questions: one True/False question (Q1), five multiple choice questions (Q2-Q6), and one with a numerical answer that had to be typed in a box (Q7). Only Q7 was not used to screen out participants.

The questions tested whether participants understood the probability of each outcome in stage 2, the minimum amount of savings in different scenarios if they had an annuity, the marginal value of tokens in different scenarios, the token-to-dollar conversion, and bonus

pay computation. Q7 tested if participants could do a simple arithmetic computation, which helped rule out simplification of the arithmetic as a mechanism for certain treatment conditions testing contingent reasoning.

If a participant failed to correctly answer one of the first six comprehension questions, the next screen would show an explanation of the correct answer and reasoning. In the case of Q2-Q6, the participant would be asked to retake the same question, and the order of the alternatives would be randomized.

To ensure good comprehension of the decision tasks in the experiment, we screened out participants with low performance in the comprehension checks, according to a set of pre-registered rules. Participants were screened out of the experiment if they failed to correctly answer more than two of the six Q1-Q6 comprehension questions on their first try, or if they failed to correctly answer a retake question. Screened-out participants were redirected to the end of the study and did not make any savings or annuity choices. Our final sample comprises only those who passed the comprehension checks and completed the study. We also retained data on whether the screened-in participants failed to answer a comprehension question correctly on their first try, and we present results about how this correlates with participants' decisions. The Supplementary Study Instructions Appendix contains the complete study instructions.[6]

1.4 Platform and Sample

Platform. The experiment was implemented through the AmeriSpeak panel from the National Opinion Research Corporation. This online panel has over 48,000 members and is designed to be representative of the U.S. household population. Households are randomly selected and heavily incentivized to participate in the panel, which reduces selection biases that can make samples unrepresentative on unobserved characteristics. On average, panelists are invited to participate in studies two to three times a month. We recruited individuals aged 18 or older.

Sample. The experiment ran from January 28, 2021 to March 4, 2021, with a pre-registered target of 3,000 participants who pass the screening questions. A total of 3,038 participants passed the screening questions and completed the study. The median duration of the study was 21.7 minutes.

AmeriSpeak collects data on financial literacy and demographic characteristics of its panel members, including educational attainment, age, gender, income, and ethnicity. Ap-

[6]Available at: https://users.nber.org/∼luttmer/StudyInstructionsAppendix.pdf

pendix Table A3 presents a summary of the demographics of our final sample, and how it compares to the U.S population. Relative to the US population, our sample is substantially more educated, but broadly similar on other demographics such as income.

2 Results

We first examine mean annuity take-up in the Benchmark conditions, and then investigate the role of psychological biases in annuity decisions by comparing take-up across experimental conditions. All standard errors are robust and clustered by participant.

Appendix Table A4 provides a complete summary of take-up in all experimental cells. Except for some of the heterogeneity analyses and the specification that pools the Salient-Contingencies treatments, all results in the figures of the body of the paper were pre-specified in the analysis plan. Appendix F details these deviations and reports the other pre-specified analyses.

2.1 Savings Decisions

Before analyzing annuity take-up, we first check whether any of the experimental conditions affected savings choices. Table 2 summarizes savings choices in all experimental conditions that were varied between participants, and shows that they are virtually identical across all conditions. This implies that any differences in annuity take-up across the conditions cannot be explained by effects of the conditions on savings.

2.2 Annuity Take-up in the Benchmark Conditions

High-price Benchmark. The first spike in Figure 2A shows that 71.4% (s.e.: 1.6) of participants in the High-price Benchmark condition bought the annuity. This leaves 28.6% of participants who didn't choose the payoff-maximizing option.

Even if savings are not chosen optimally, the expected payout is weakly higher with the annuity. Buying the annuity keeps expected payoffs constant only if a participant saves optimally without an annuity and makes the payoff-*minimizing* savings choice with the annuity. This occurs for 14.1% of participants, but take-up for the remaining 85.9% is only 72.0%. Even among the 78.3% for whom the annuity was strictly stochastically dominant, take-up is 73.6%. Hence, suboptimal savings choices do not explain the lack of annuity take-up.

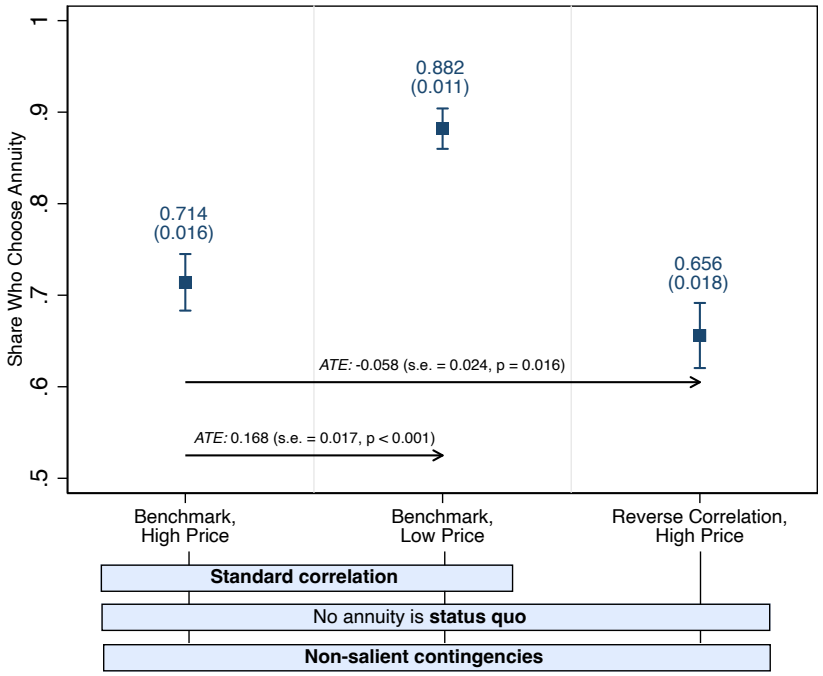
Table 2: Mean savings by annuity condition

	Number of Participants	No annuity (Optimal savings = 40)	Low-price annuity (Optimal savings = 10)	High-price annuity (Optimal savings = 10)
Panel A. Full sample, of which:	3038	43.24 (0.08)	25.75 (0.22)	20.30 (0.15)
Savings first	780	43.57 (0.15)	26.63 (0.43)	21.02 (0.30)
Savings second, benchmark (regular correlation, status quo)	822	43.01 (0.15)	24.40 (0.42)	19.33 (0.29)
Savings second, no status quo (regular correlation)	750	43.37 (0.16)	27.12 (0.44)	21.37 (0.31)
Savings second, reverse correlation (status quo, "insurance" wording)	686	42.99 (0.16)	24.85 (0.47)	19.49 (0.33)
Panel B. Sample with regular correlation, of which:	2352	43.31 (0.09)	26.01 (0.25)	20.54 (0.17)
Savings first	780	43.57 (0.15)	26.63 (0.43)	21.02 (0.30)
Savings second	1572	43.18 (0.11)	25.70 (0.30)	20.30 (0.21)
Panel C. Sample with regular correlation, of which:	2352	43.31 (0.09)	26.01 (0.25)	20.54 (0.17)
"Annuity" wording	793	43.40 (0.15)	26.20 (0.42)	20.98 (0.30)
"Social Security" wording	762	43.10 (0.15)	25.52 (0.44)	20.21 (0.31)
"Insurance" wording	797	43.41 (0.15)	26.29 (0.44)	20.42 (0.30)
Panel D. Sample with savings second, status quo, and "insurance" wording, of which:	971	43.09 (0.14)	24.97 (0.40)	19.53 (0.28)
Benchmark (regular correlation), "insurance" wording only	285	43.31 (0.26)	25.26 (0.73)	19.65 (0.50)
Reverse correlation ("insurance" wording)	686	42.99 (0.16)	24.85 (0.47)	19.49 (0.33)

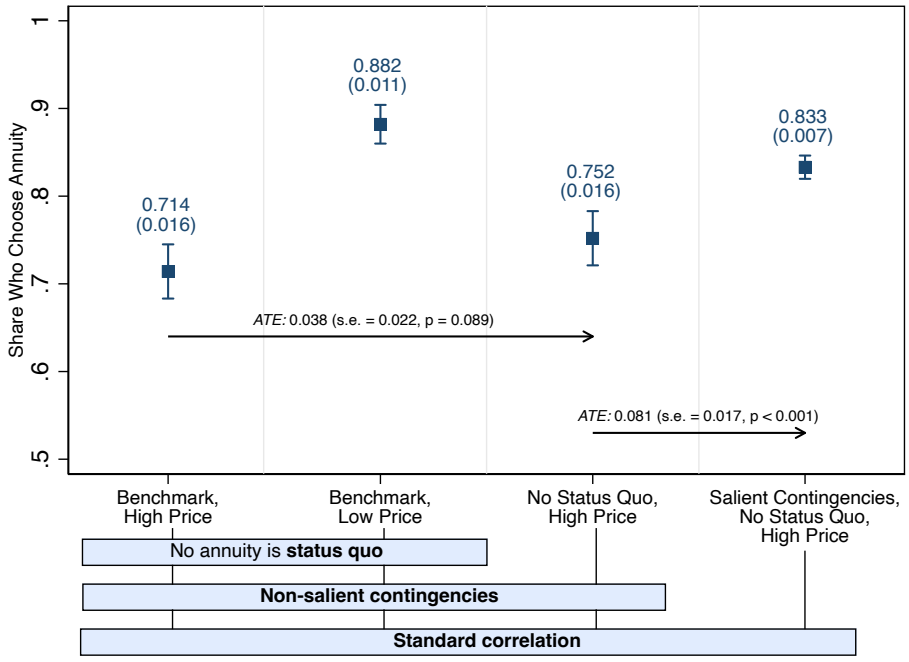
Notes: These panels report the mean savings by the specified experimental conditions. Panel A contains a summary of the full sample, while panels B, C, and D focus on the effect of the savings order, effect of wording used, and effect of reverse correlation respectively. All standard errors are clustered at the participant level.

Figure 2: Share choosing annuity by treatment group

(a) Reverse correlation



(b) Salient contingencies



Notes: The spikes in this figure show the share of participants in each group who took up the annuity. The arrows indicate average treatment effects (ATEs) of the experimental treatments under the arrowheads relative to the experimental conditions under the beginning of the arrows. Above each spike is the mean take-up within the group (indicated by the marker), with the standard error in parentheses. The vertical lines in the spikes represent 95% confidence intervals. All standard errors are clustered at the participant level.

Low-price Benchmark. In the Low-price Benchmark condition, the price of the annuity is 10 tokens rather than 20 tokens. Optimal savings remain unchanged, but a participant making optimal savings decisions now earns $7.50 when buying the annuity.

The second spike of Figure 2A shows that annuity take-up increased to 88.2% (s.e.: 1.1) at the lower price. This 16.8 (s.e. 1.7) percentage-point change shows that at least 58.7% of participants who declined to buy the high-price annuity did not do so out of some immutable unwillingness to buy annuities or due to disengagement from the experiment. Rather, the price elasticity is consistent with participants understanding the experiment and appreciating that the annuity has positive value, but having a negative bias in how they construe its value.

2.3 Reverse Correlation—Receiving a Contingent Payment in the Low-payout State

The third spike of Figure 2A shows that the Reverse-Correlation treatment reduced take-up by 5.8 (s.e.: 2.4) percentage points. This finding rejects the hypothesis that a reluctance to buy state-contingent contracts that pay out in "good" states contributes to low annuity take-up. This rejection implies that the incomplete take-up of state-contingent contracts is not specific to annuities, but also applies to insurance more generally. Failures of contingent reasoning may thus be relevant not just to annuities but also to insurance take-up in environments where people can both buy insurance and self-ensure through precautionary savings.

Impact of "insurance" wording. Because the Reverse-Correlation condition always used "insurance" wording to describe the annuity, it is important to check if some of the effect of this treatment could be due to mere word choice. The analysis plan specified that if, in the High-price Benchmark condition, take-up for the "insurance" wording was not significantly different from the other two wordings ("annuity" or "Social Security") at the 10% level, the Reverse-Correlation condition would be compared to all three wordings in the Benchmark condition, as we have done above. The p-value for this is 0.571, implying that wording choice did not significantly affect participant decisions. Appendix Table A4 presents take-up by wording in all conditions, showing that it has no systematic effect on outcomes. The participants' lack of reaction to choice of wording indicates that framing manipulations do not easily influence their annuity or savings decisions in the experiment.

2.4 Failures to Reason Through Contingencies

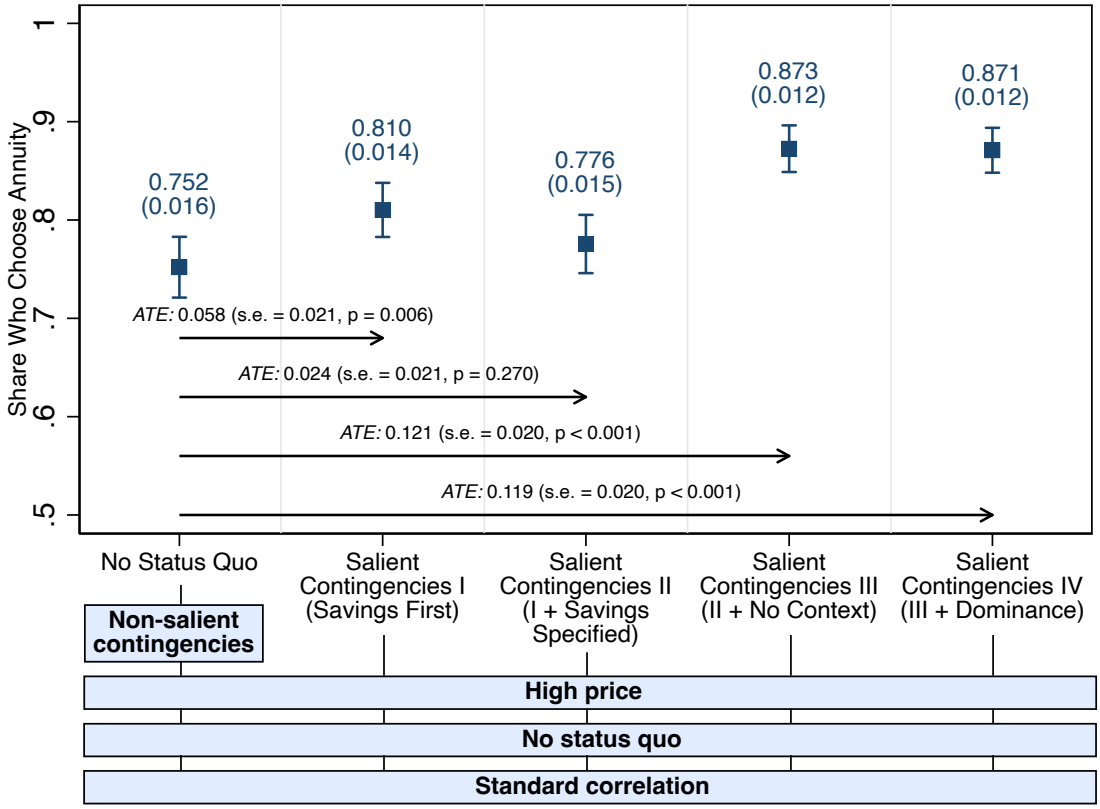
Overall effect of Salient-Contingencies manipulations. The first two spikes of Figure 2B show the two Benchmark cases as reference. The third spike shows that removing the status quo increased take-up by 3.8 (s.e.: 2.2) percentage points. This increase is only marginally significant (p-value: 0.089), but qualitatively consistent with Brown et al. (2017) and Brown et al. (2021). The fourth spike shows that 83.3% (s.e.: 0.7) chose the annuity when contingencies were made salient through any of the four Salient-Contingencies treatments. The 8.1 (se.: 1.7) percentage-point increase over the No-Status-Quo condition equals 48.2% of the effect size of lowering the price in the Benchmark conditions, indicating that failures of contingent reasoning are a meaningful impediment to annuity take-up.

Effects by type of Salient-Contingencies manipulation. Figure 3 examines take-up separately by each variant of the Salient-Contingencies manipulations, and again compares these to take-up in the No-Status-Quo condition. The second spike shows take-up for Salient Contingencies I, where the annuity decision came after the three savings questions rather than at the very start of the decision tasks. This manipulation increased annuity take-up by 5.8 (s.e.: 2.1) percentage points. Thus, participants have a greater appreciation of the annuity's value if they are required to first think about how savings affect payouts with and without the annuity.

In Salient Contingencies II, we specified the participant's prior savings choices with and without the annuity. The third spike shows that this manipulation insignificantly (p-value 0.270) increased take-up relative to the No-Status-Quo condition, but decreased it relatively to Salient Contingencies I. Because this treatment instructed participants how much they had to save with and without the annuity (without reminding them that these were their own choices), participants may have perceived a loss of autonomy. As Bartling et al. (2014) show, people value autonomy in decision-making. Hence, the perceived loss of autonomy may have led them to disengage with the experiment and/or attenuate their perception of the incremental value of annuity option.

In Salient Contingencies III, we kept the diagram the same as in Salient Contingencies II, but the introductory text no longer described the tokens as coming from income, savings, or an annuity. This lack of context removed the potential perceived loss of autonomy over savings decisions. The fourth spike shows that this condition increased take-up to 87.3% (s.e.: 1.2), which nearly matches the Low-price Benchmark take-up rate of 88.2% (s.e.: 1.1). Relative to the No-Status-Quo condition, this implies a treatment effect of 12.1 (s.e.: 2.0) percentage points.

Figure 3: Share choosing annuity by salient-contingencies condition



Notes: The spikes in this figure show the share of participants in each group who took up the annuity. The arrows indicate average treatment effects (ATEs) of the experimental treatments under the arrowheads relative to the experimental conditions under the beginning of the arrows. Above each spike is the mean take-up within the group (indicated by the marker), with the standard error in parentheses. The vertical lines in the spikes represent 95% confidence intervals. All standard errors are clustered at the participant level.

The effect of Salient Contingencies II and III is unlikely to be due to simpler arithmetic calculations, as participants had easy access to an online calculator displayed on their decision-making screen. Moreover, for the 82.7% of participants who correctly answered the comprehension question consisting of an arithmetic calculation analogous to the types required in the experiment (Q7), the treatment effects of Salient Contingencies II and III relative to the No-Status-Quo condition were similar, at 3.7 (s.e.: 2.4) and 12.1 (s.e.: 2.2) percentage points, respectively. Rather, these treatment conditions made it more transparent how the purchase of the annuity interacts with savings choices and ultimate payouts.

Salient Contingencies IV presents the annuity choice in exactly the same way as in Salient Contingencies III, but alters the saving level corresponding to the annuity to ensure that it is strictly stochastically dominant and that the number of tokens in stage 2 is the same as without the annuity. The fifth spike shows that Salient Contingencies IV results in an 87.1%

(s.e.: 1.2) take-up, which is almost identical to take-up without the adjustment. Moreover, if we limit the sample to participants for whom we needed to adjust the savings rate to ensure stochastic dominance, the difference in take-up rates between Salient Contingencies III and IV remains statistically indistinguishable from zero (see Appendix Table A5). The lack of an effect of the dominance adjustment implies that small-stakes risk aversion does not play a meaningful role in annuity take-up in our experiment.

Salient Contingencies IV present an extremely simple choice: the screen shows that the annuity option provides strictly more tokens in stage 1 and the same number of tokens in stage 2. Hence, as long as a participant understands that more tokens correspond to more dollars, they would not even need to know how tokens convert to dollars to determine that the annuity provides strictly higher payoffs whether or not they survive. The 13% of participants who selected a transparently stochastically dominated option in a context-free environment were likely participants who had disengaged from the experiment.

Finally, we find that Salient Contingencies IV results in a take-up of 87.5% (s.e.: 1.6%) percent in the Reverse-Correlation condition, which is 21.9 (s.e.: 2.4) percentage points higher than in the Reverse-Correlation condition with status quo and without salient contingencies (Appendix Table A4). This suggests that failures of contingent reasoning also affect regular insurance decisions.

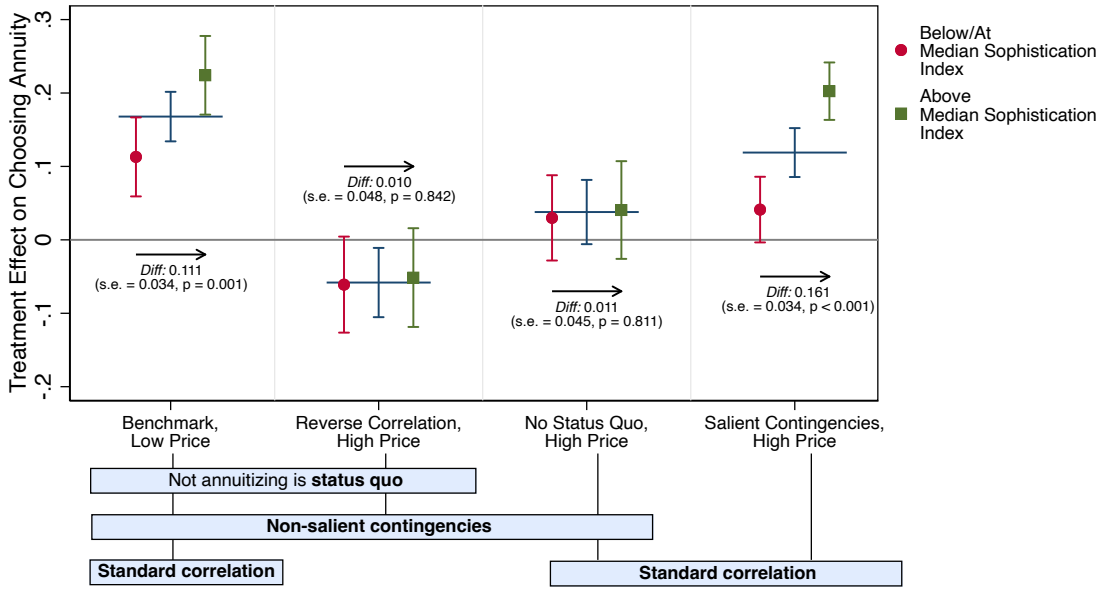
2.5 Heterogeneity by Measures of Decision-Making Sophistication

Figure 4A shows annuity take-up rates for the sample as a whole (the horizontal line with a confidence interval), and separately for three measures of decision-making sophistication: answering the three standard financial literacy questions (Lusardi and Mitchell (2011)) correctly (diamonds and squares), selecting payoff-maximizing savings choices (pluses and crosses), and answering all comprehension questions correctly in the first attempt (circles and triangles). Responses by those with lower levels of decision-making sophistication are shown on the left in orange-reddish colors, and those with higher levels of sophistication are on the right in blue-greenish colors. The panel shows that more sophisticated participants react more strongly than less sophisticated ones to the two treatments that elicit a strong response in the sample as a whole: changing the price of the annuity and making contingencies salient.

Figure 4: Annuity take-up and treatment effects by decision-making sophistication

(a) Take-up means by decision-making sophistication

(b) Treatment effects by decision-making sophistication



Notes: Panel (a) shows the share of participants who took up the annuity by experimental group and by proxies for decision-making sophistication. Panel (b) shows the treatment effect on annuity take-up by an index for decision-making sophistication, constructed by standardizing the three comprehension proxies and taking their mean. The treatment effects for the first three groups are relative to the High-price Benchmark group. The treatment effect of the fourth group (Salient Contingencies, High Price) are relative to the No-Status-Quo High-price group. The text below the arrows reports the difference in treatment effects between participants with above- versus below-median values of the sophistication index. In all panels, the vertical spikes indicate the 95% confidence interval and standard errors are clustered at the participant level.

Because all three measures of decision-making sophistication show the same pattern, we combine the underlying values (i.e., fraction of financial literacy questions answered correctly, fraction of payoff-maximizing savings choices made, and fraction of comprehension questions answered correctly) into a single index by standardizing the variables and taking their average. The red circles and green squares in Figure 4B show treatment effects by those with below- and above-median sophistication, respectively, while the horizontal lines show average treatment effects. Treatment effects by each of the three measures separately are shown in Appendix Figure A2. Reducing the price of the annuity by half causes more-sophisticated participants to increase their annuity take-up by about twice as much as less-sophisticated ones. The Salient-Contingencies treatment elicits an increase in take-up among more-sophisticated participants that is four times as large as among less-sophisticated ones.

The finding that the most sophisticated respondents reacted the most to the Salient-Contingencies treatments indicates that these treatment effects were not an artifact of respondent confusion. Rather, it suggests that reasoning through contingencies is cognitively challenging, including for the most-sophisticated individuals. A possible explanation of why less-sophisticated participants exhibited smaller treatment effects is that their choices are driven by automatic heuristics that are not necessarily taking into account key features of the alternatives, like the specified price or the displayed payoffs.

Appendix Figure A1 presents the impact of each type of Salient-Contingencies treatment by the sophistication index. More-sophisticated participants significantly increase their take-up in response to each type of treatment. However, the first two Salient-Contingencies treatments have zero or even negative effects on the take-up of less-sophisticated participants. Salient-Contingencies treatments III and IV increase less-sophisticated participants' take-up, but to a lesser degree than that of the more-sophisticated participants. Appendix Figure A3 replicates Figure 4B for financial literacy, education, income, and age. Appendix tables A6 and A7 provide heterogeneity analysis by sample cuts around the median in number of correct answers to the financial literacy questions, income, educational attainment, and age. There is significant heterogeneity in treatment effects of salient contingencies by education and financial literacy—with participants who are better educated and more financially literate reacting more strongly—but not by age or income. These findings are consistent with the heterogeneity results by decision-making sophistication.

2.6 Interpreting Effect Sizes

The benchmark annuitization rate in our experiment, 71.4%, is much larger than the estimates of actual annuitization rates in the field. For example, a recent survey shows that only

12% of Americans over age 50 with investable assets of more than $\$100,000$ own an annuity (Arapakis and Wettstein, 2023). The substantially larger take-up in our experiment is by design. First, we eliminated all rational reasons for not annuitizing, such as bequest motives or the need for liquidity in late-in-life uninsurable shocks (e.g., related to long-term care needs). Second, we presented participants with a binary choice to annuitize or not, which means that even if participants were choosing at random, annuity take-up would be 50%. By contrast, outside the lab, most people are rarely presented with a direct choice to annuitize, and many people may not even be aware that such an option exists. Naturally, the relatively high benchmark rate of annuitization limits the degree to which our experimental treatments can further increase annuitization, and makes it difficult to extrapolate our estimated effects on take-up to settings outside our controlled experiment.

An arguably more portable metric is to express treatment effect magnitudes in terms of their relative impact on perceived value. We do this with a simple model where a person i in condition j takes up the annuity if $\tilde{v}_j - p_j + \varepsilon_{ij} > 0$, where $\tilde{v}_j$ is the average perceived value of the annuity in experimental condition j, p_j is the price, and $\varepsilon_{ij} \sim N(0, \sigma^2)$ is either experimental decision noise or idiosyncratic differences in perceptions of value. Thus, choice follows a Probit model, and we think of the non-price experimental variation as altering the perceived value $\tilde{v}$. For a given experimental treatment j, we therefore model the treatment effect τ_j as changing the average baseline valuation $\tilde{v}_0$ (in the benchmark condition) to $\tilde{v}_j := \tilde{v}_0 + \tau_j$. This model formalizes the intuition that the impact of any given treatment τ_j on take-up will decrease with $\tilde{v}_0$ (for $\tilde{v}_0 \geq 0$), so that high values of $\tilde{v}_0$ will generate "ceiling effects" where treatments that have a relatively large impact on the perceived value of an annuity might still have a small impact on take-up. Because take-up of annuities is high in our experiment, it is therefore helpful to quantify treatment effects in terms of their impact on perceived annuity value.

Through the lens of this model, consider annuity take-up in the benchmark condition, where take-up is 71.4% or 88.2% when the price is 20 or 10 tokens, respectively. The model parameters $\tilde{v}_0$ and σ^2 are identified from these take-up rates, as the model implies that $\Phi((\tilde{v}_0 - 20)/\sigma) = 0.714$ and $\Phi((\tilde{v}_0 - 10)/\sigma) = 0.882$, where Φ is the standard normal CDF. We estimate that $\tilde{v}_0$ in the benchmark condition is 29.1, meaning that the price would need to rise to 29.1 tokens for annuity take-up to decrease to 50%. Appendix E presents the estimates of the Probit model and corresponding average valuations.

With these parameters in hand, we rescale the treatment effects to be in units of equivalent price changes, or marginal utility. Removing the status quo increased take-up by 3.8 percentage points, which in our model is equivalent to reducing the price by 1.9 tokens, or 6.4% of the average perceived annuity value $\tilde{v}_0$. Inducing people to think through their

savings choices prior to their annuity decision (Salient Contingencies I) increased take-up by 5.8 percentage points relative to the no status quo group, which is equivalent to a 3.2 token decrease in the price, or 11.0% of the average perceived annuity value $\tilde{v}_0$. Finally, removing all context in the Salient Contingencies III treatment increased take-up by 12.1 percentage points relative to the no status quo group, which is equivalent to a 7.4 token decrease in the price, or 25.4% of the average perceived annuity value $\tilde{v}_0$. That is, if people did not exhibit failures of contingent reasoning, our results imply that their valuation of annuities would be 25.4% higher. Given that take-up of the annuity is utility-maximizing for everyone in our experiment, these rescaled estimates of our treatment effects suggest a large role for biases.

3 Concluding Remarks

In a tightly controlled experiment, we find that take-up of annuities increases in response to treatments that reduce failures of contingent reasoning. However, we reject the hypothesis that people find "longevity insurance" less natural than insurance for a "bad" state of the world. When participants must purchase insurance against losing income, with the financial consequences of that insurance identical to the financial consequences of purchasing an annuity, participants still substantially under-insure and still display systematic failures of contingent reasoning. Thus, our results suggest that failures of contingent reasoning may also lower take-up of insurance in other domains where people can self-insure through precautionary savings.

Our experiment was intentionally simple to generate an unambiguous benchmark for optimal choice and to cleanly elucidate psychological barriers to take-up. While this is novel relative to the existing literature—where it is not fully clear whether choosing an annuity is optimal or not—an important limitation of our approach is that the choices participants are presented with are more stylized than the choices they face in reality. Thus, the exact magnitudes that we estimate are unlikely to generalize.

But we are more confident in our *qualitative* results. First, these results are unlikely to be artifacts of participant confusion or experimenter demand effects. Our protocol excluded participants who did not answer the majority of the comprehension checks correctly, and among the remaining participants those with higher levels of comprehension were, if anything, more likely to exhibit failures of contingent reasoning. We also tried to minimize demand effects by carefully designing the experiment to avoid signaling to participants that one or the other choice was better, and the lack of any treatment effects of the language we used (see Section 2.3) suggests that participants were not attempting to infer which choice is better based on the language we used.

Second, while there may be concern that the low stakes in our setting are conducive of mistakes, a counteracting force is that participants were provided with simple and clear instructions that enabled them to make optimal choices. While the stakes are higher in people's actual annuity decisions in the field, reasoning through contingencies is also much more complex because in practice people must consider many more contingencies, and typically with less guidance. Moreover, meta-analyses of the importance of stakes in economics experiments have not found meaningful effects for systematic behavioral biases in reasoning (Camerer and Hogarth, 1999; Enke et al., 2023). Thus, it is well possible that failures of contingent reasoning may be even more pronounced in field settings.

Many other forms of systematic biases in Behavioral Economics have first been established in controlled laboratory settings with lower stakes, and later explored in field settings. We see our study as a necessary first step in this process. A natural next step might be to examine correlations between mistakes in our study and people's revealed preferences over the kinds of annuity products that are actually offered by the market. Another natural next step might be to implement elements of our treatments in the field. For example, paralleling our first salient contingencies treatment, it might be interesting to ask people to first think through the savings decisions they would make with and without an annuity (perhaps using a guided planning exercise), and only then to ask them to make an annuity decision.

Continued research along these lines can provide valuable further guidance for policies that influence annuity choice. The finding that failures of contingent reasoning reduce annuity take-up implies that private annuity decisions are suboptimal and that therefore that on the margin, policies increasing annuity take-up are welfare improving. A second implication of our findings is that policies that make it easier to think through contingencies, for example by spelling out key contingencies, or that reduce the number of relevant contingencies, for example by simplifying the annuity, will bring annuity demand closer to the welfare-maximizing level. Of course, all of these potential implications need to be fleshed out in formal economic models, as our experimental results are partial equilibrium effects, and we do not consider how annuity prices might adjust to interventions that reduce biases in their take-up. Continued research along these lines can provide valuable guidance for improving annuity markets.

References

Agnew, Julie R, Lisa R Anderson, Jeffrey R Gerlach, and Lisa R Szykman, "Who Chooses Annuities? An Experimental Investigation of the Role of Gender, Framing, and Defaults," *American Economic Review: Papers & Proceedings*, 2008, *98* (2), 418–22.

Ambuehl, Sandro, B Douglas Bernheim, and Annamaria Lusardi, "Evaluating deliberative competence: A simple method with an application to financial choice," *American Economic Review*, 2022, *112* (11), 3584–3626.

Ameriks, John, Andrew Caplin, Steven Laufer, and Stijn Van Nieuwerburgh, "The Joy of Giving or Assisted Living? Using Strategic Surveys to Separate Public Care Aversion from Bequest Motives," *Journal of Finance*, 2011, *66* (2), 519–561.

Arapakis, Karolos and Gal Wettstein, "How much do people value annuities and their added features?," *Center for Retirement Research Working Paper 2023-18*, 2023.

Bartling, Björn, Ernst Fehr, and Holger Herz, "The Intrinsic Value of Decision Rights," *Econometrica*, 2014, *82* (6), 2005–2039.

Bernheim, B Douglas and Dmitry Taubinsky, "Behavioral Public Economics," in B Douglas Bernheim, Stefano DellaVigna, and David Laibson, eds., *The Handbook of Behavioral Economics*, Vol. 1, New York: Elsevier, 2018.

Beshears, John, James Choi, David Laibson, and Brigitte Madrian, "Behavioral Household Finance," in B. Douglas Bernheim, Stefano DellaVigna, and David Laibson, eds., *The Handbook of Behavioral Economics*, Vol. 1, New York: Elsevier, 2018.

_ , James J Choi, David Laibson, Brigitte C Madrian, and Stephen P Zeldes, "What Makes Annuitization More Appealing?," *Journal of Public Economics*, 2014, *116*, 2–16.

Brown, Jeffrey R, "Rational and Behavioral Perspectives on the Role of Annuities in Retirement Planning," in Annamaria Lusardi, ed., *Overcoming the Saving Slump: How to Increase the Effectiveness of Financial Education and Saving Programs*, Chicago: University of Chicago Press, 2009, pp. 178–206.

_ , Arie Kapteyn, and Olivia S Mitchell, "Framing and Claiming: How Information Framing Affects Expected Social Security Claiming Behavior," *Journal of Risk and Insurance*, 2016, *83* (1), 139–162.

_ , _ , Erzo FP Luttmer, and Olivia S Mitchell, "Cognitive Constraints on Valuing Annuities," *Journal of the European Economic Association*, 2017, *15* (2), 429–462.

_ , _ , _ , _ , and Anya Samek, "Behavioral Impediments to Valuing Annuities: Complexity and Choice Bracketing," *Review of Economics and Statistics*, 2021, *103* (3), 533–546.

_ , Jeffrey R Kling, Sendhil Mullainathan, and Marian V Wrobel, "Why Don't People Insure Late-Life Consumption? A Framing Explanation of the Under-Annuitization Puzzle," *American Economic Review*, 2008, *98* (2), 304–09.

_ , _ , _ , and _ , "Framing Lifetime Income," *Journal of Retirement*, 2013, *1* (1), 27–37.

Camerer, Colin F. and Robin M. Hogarth, "The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework," *Journal of Risk and Uncertainty*, 1999, *19*, 7–42.

Chakraborty, Anujit and Chad W Kendall, "Noisy Foresight," *NBER Working Paper*, 2022.

Davidoff, Thomas, Jeffrey R Brown, and Peter A Diamond, "Annuities and Individual Welfare," *American Economic Review*, 2005, *95* (5), 1573–1590.

Enke, Benjamin, Uri Gneezy, Brian Hall, David Martin, Vadim Nelidov, Theo Offerman, and Jeroen van de Ven, "Cognitive Biases: Mistakes or Missing Stakes?," *Review of Econoimcs and Statistics*, 2023, *105* (4), 818–832.

Esponda, Ignacio and Emanuel Vespa, "Hypothetical Thinking and Information Extraction in the Laboratory," *American Economic Journal: Microeconomics*, 2014, *6* (4), 180–202.

_ and _ , "Contingent Thinking and the Sure-Thing Principle: Revisiting Classic Anomalies in the Laboratory," *Review of Economic Studies*, 2024, *91* (5), 2806–2831.

Gazzale, Robert S and Lina Walker, "Behavioral Biases in Annuity Choice: An Experiment," *Williams College Economics Department Working Paper Series*, 2009.

Li, Shengwu, "Obviously Strategy-Proof Mechanisms," *American Economic Review*, 2017, *107* (11), 3257–87.

Lockwood, Lee M, "Bequest Motives and the Annuity Puzzle," *Review of Economic Dynamics*, 2012, *15* (2), 226–243.

Lusardi, Annamaria and Olivia S Mitchell, "Financial Literacy around the World: An Overview," *Journal of Pension Economics & Finance*, 2011, *10* (4), 497–508.

Martínez-Marquina, Alejandro, Muriel Niederle, and Emanuel Vespa, "Failures in Contingent Reasoning: The Role of Uncertainty," *American Economic Review*, October 2019, *109* (10), 3437–74.

Mitchell, Olivia S, John Piggott, and Noriyuke Takayama, *Revisiting Retirement Payouts: Market Developments and Policy Issues*, Oxford, Oxford University Press, 2011.

Peijnenburg, Kim, Theo Nijman, and Bas JM Werker, "Health Cost Risk: A Potential Solution to the Annuity Puzzle," *Economic Journal*, 2017, *127* (603), 1598–1625.

Poterba, James, Steven Venti, and David Wise, "The Composition and Drawdown of Wealth in Retirement," *Journal of Economic Perspectives*, 2011, *25* (4), 95–118.

Rabin, Matthew and Georg Weizsäcker, "Narrow Bracketing and Dominated Choices," *American Economic Review*, September 2009, *99* (4), 1508–43.

Read, Daniel, George Loewenstein, and Matthew Rabin, "Choice Bracketing," *Journal of Risk and Uncertainty*, 1999, *19*, 171–197.

Reichling, Felix and Kent Smetters, "Optimal Annuitization with Stochastic Mortality and Correlated Medical Costs," *American Economic Review*, 2015, *105* (11), 3273–3320.

Shafir, Eldar and Amos Tversky, "Thinking through Uncertainty: Nonconsequential Reasoning and Choice," *Cognitive psychology*, 1992, *24* (4), 449–474.

Tversky, Amos and Daniel Kahneman, "The Framing of Decisions and the Psychology of Choice," *Science*, 1981, *211* (4481), 453–458.

Yaari, Menahem E, "Uncertain Lifetime, Life Insurance, and the Theory of the Consumer," *Review of Economic Studies*, 1965, *32* (2), 137–150.

Online Appendix

Failures of Contingent Reasoning in Annuitization Decisions

Erzo F.P. Luttmer ⓡ Priscila de Oliveira ⓡ Dmitry Taubinsky

Table of Contents

A Additional Experimental Details

Table A1: Financial incentives

	N	Expected payout		
		With annuity (I)	Without annuity (II)	Incentive to annuitize (I) - (II)
High-Price Benchmark	822	3.83 (1.04)	2.12 (0.52)	1.71 (1.00)
Low-Price Benchmark	822	5.70 (1.50)	2.12 (0.52)	3.58 (1.44)
High-Price No Status Quo	750	3.58 (1.05)	2.08 (0.54)	1.50 (0.95)
High-Price Salient Contingencies I	780	3.62 (1.05)	2.05 (0.54)	1.57 (0.97)
High-Price Salient Contingencies II	762	3.70 (1.05)	2.11 (0.52)	1.58 (0.99)
High-Price Salient Contingencies III	761	3.69 (1.07)	2.10 (0.53)	1.59 (0.98)
High-Price Salient Contingencies IV	829	4.56 (0.55)	2.06 (0.55)	2.50 (0.00)
High-Price Reverse Correlation	686	8.81 (1.08)	7.13 (0.52)	1.69 (1.02)
Low-Price Reverse Correlation Salient Contingencies IV	686	12.13 (0.52)	7.13 (0.52)	5.00 (0.00)

Notes: This table reports the average expected payout for respondents by treatment conditions. Each row corresponds to a different treatment condition. Column 1 reports the average payout from the annuity option. Column 2 reports the average payout of the option that does not include an annuity. Column 3 reports the difference between columns 1 and 2, which represents the average monetary gain from choosing the annuity option. Standard deviations in parentheses. The Salient Contingencies IV treatment involves adjusting savings levels so that the annuity is stochastically dominant. As a result, the incentive for choosing the annuity is exactly equal to $2.50 for the high-price annuity and exactly equal to $5.00 for the low-price annuity.

Table A2: Order of questions

Expected Fraction	Nr. of People	Savings First	Annuity Block 1		Savings Second	Annuity Block 2		Annuity Block 3 (Added partway through)
Panel A: Benchmark								
1/8	426		Low-price Benchmark	High-price Benchmark	X	High-price Salient Contingencies II or III or IV		Low-price Salient Contingencies II or III or IV
1/8	396		High-price Benchmark	Low-price Benchmark	X	High-price Salient Contingencies II or III or IV		Low-price Salient Contingencies II or III or IV
Panel B: No Status Quo								
1/8	380		Low-price No Status Quo	High-price No Status Quo	X	High-price Salient Contingencies II or III or IV		Low-price Salient Contingencies II or III or IV
1/8	370		High-price No Status Quo	Low-price No Status Quo	X	High-price Salient Contingencies II or III or IV		Low-price Salient Contingencies II or III or IV
Panel C: Salient Contingencies								
1/16	209	X	Low-price Salient Contingencies I	High-price Salient Contingencies I		High-price Salient Contingencies II or III or IV		Low-price Salient Contingencies II or III or IV
1/16	193	X	High-price Salient Contingencies I	Low-price Salient Contingencies I		High-price Salient Contingencies II or III or IV		Low-price Salient Contingencies II or III or IV
1/16	195	X	High-price Salient Contingencies II or III or IV			Low-price Salient Contingencies I	High-price Salient Contingencies I	Low-price Salient Contingencies II or III or IV
1/16	183	X	High-price Salient Contingencies II or III or IV			High-price Salient Contingencies I	Low-price Salient Contingencies I	Low-price Salient Contingencies II or III or IV
Panel D: Reverse Correlation								
1/8	325		Low-price Reverse-Correlation	High-price Reverse-Correlation	X	Low-price Reverse-Correlation Salient Contingencies IV		High-price Reverse-Correlation Salient Contingencies IV
1/8	361		High-price Reverse Correlation	Low-price Reverse Correlation	X	Low-price Reverse-Correlation Salient Contingencies IV		High-price Reverse-Correlation Salient Contingencies IV

Notes: This table describes all of the different experimental cells for annuity take-up decisions. Cells in gray were not included in the pre-analysis plan nor in the analyses in the body of the paper. Additionally, the three possible wordings ("annuity," "Social Security," "insurance") were randomized at the participant level, except for the reverse-correlation treatments, which used only the "insurance" wording.

B Demographics

Table A3: Demographic characteristics

	Experimental Sample	U.S. Adult Population
Female	0.54	0.52
Age (median)	54.0	47.0
Bachelor's degree or higher	0.53	0.31
Employed	0.65	0.63
Household income ($, median)	67500	78040
Non-Hispanic White	0.77	0.65
Non-Hispanic Black	0.06	0.13
Hispanic	0.09	0.16
Married	0.58	0.51
Financial literacy I (interest)	0.92	-
Financial literacy II (inflation)	0.85	-
Financial literacy III (risk exposure)	0.91	-

Notes: Column 1 of this table reports means (unless stated otherwise) for various demographic variables for the 3,038 participants who completed the study. The second column reports the statistics for the U.S. adult population living in households from the 2019 American Community Survey 1-Year Estimates Public Use Microdata Sample. The variable *Financial literacy* is an indicator for whether the participant answered the following three questions from Lusardi and Mitchell (2011) correctly: "Suppose you had $100 in a savings account and the interest rate was 2% per year. After 5 years, how much do you think you would have in the account if you left the money to grow: more than $102, exactly $102, or less than $102?" (which corresponds to Financial literacy I in the table above), "Imagine that the interest rate on your savings account was 1% per year and inflation was 2% per year. After 1 year, with the money in this account would you be able to buy: more than, exactly the same as, or less than today?" (Financial literacy II) and "Do you think that the following statement is true or false? Buying a single company stock usually provides a safer return than a stock mutual fund" (Financial literacy III).

C Annuity Take-Up for All Treatment Cells

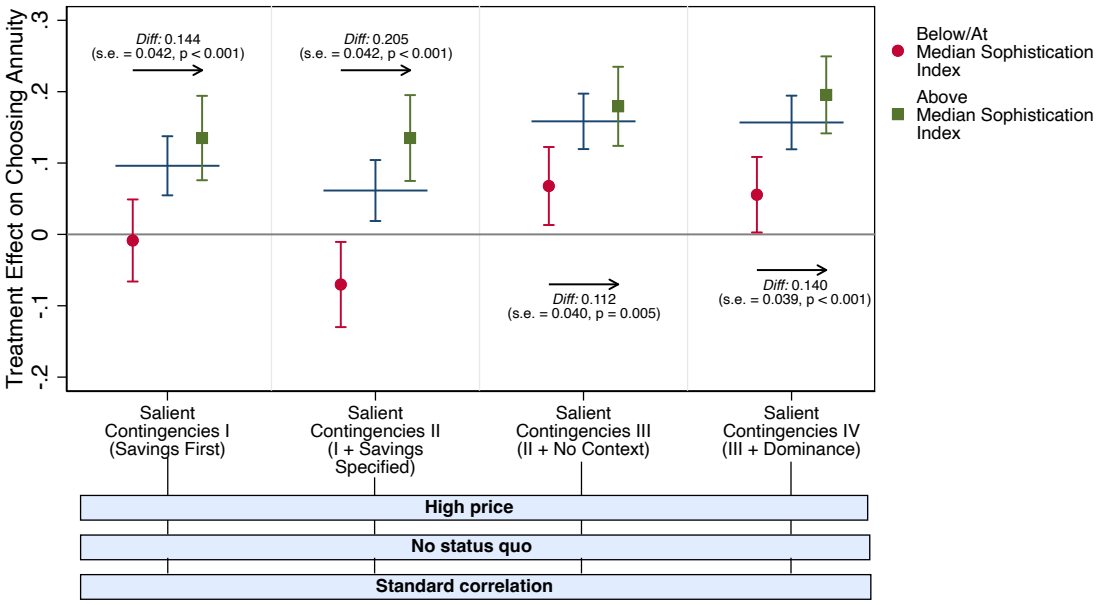
Table A4: Annuity take-up means for all treatment cells

	N	Take-up mean by wording used			
		All wordings	"Annuity"	"Social Security"	"Insurance"
Panel A. Benchmark (has status quo, regular correlation)					
High price	822	0.714 (0.016)	0.660 (0.029)	0.752 (0.026)	0.726 (0.026)
Low price	822	0.882 (0.011)	0.873 (0.021)	0.903 (0.018)	0.870 (0.020)
Panel B. No Status Quo (has regular correlation)					
High price	750	0.752 (0.016)	0.738 (0.029)	0.749 (0.028)	0.767 (0.026)
Low price	750	0.827 (0.014)	0.797 (0.026)	0.831 (0.024)	0.848 (0.022)
Panel C. Reverse Correlation (only has "insurance" wording)					
Status Quo, non-salient contingencies					
High price	686	n/a n/a	n/a n/a	n/a n/a	0.656 (0.018)
Low price	686	n/a n/a	n/a n/a	n/a n/a	0.810 (0.015)
No status quo, Salient Contingencies IV (III + dominance)					
High price	432	n/a n/a	n/a n/a	n/a n/a	0.875 (0.016)
Low price	686	n/a n/a	n/a n/a	n/a n/a	0.854 (0.013)
Panel D. Salient Contingencies (has no status quo, regular correlation)					
High price					
Salient Contingencies I, (savings first)	780	0.810 (0.014)	0.805 (0.023)	0.801 (0.026)	0.826 (0.024)
Salient Contingencies II, (I + savings specified)	762	0.776 (0.015)	0.765 (0.026)	0.786 (0.027)	0.777 (0.026)
Salient Contingencies III, (II + no context)	761	0.873 (0.012)	0.888 (0.020)	0.866 (0.021)	0.863 (0.022)
Salient Contingencies IV, (III + dominance)	829	0.871 (0.012)	0.878 (0.020)	0.891 (0.019)	0.845 (0.022)

	N	Take-up mean by wording used			
		All wordings	"Annuity"	"Social Security"	"Insurance"
Panel D (continued). Salient contingencies (has no status quo, regular correlation)					
Low price					
Salient Contingencies I, (savings first)	780	0.864 (0.012)	0.869 (0.020)	0.838 (0.024)	0.884 (0.021)
Salient Contingencies II, (I + savings specified)	487	0.877 (0.015)	0.864 (0.027)	0.887 (0.026)	0.879 (0.025)
Salient Contingencies III, (II + no context)	528	0.879 (0.014)	0.868 (0.026)	0.891 (0.023)	0.876 (0.025)
Salient Contingencies IV, (III + dominance)	542	0.893 (0.013)	0.892 (0.024)	0.908 (0.021)	0.880 (0.024)

Notes: This table reports the means of annuity take-up by wording used and by the specified treatments. Rows in gray were not included in the pre-analysis plan. Panel A displays the results for the Benchmark groups; panel B displays the results for the groups in which there is no status quo in the annuity choice; panel C displays the results for the groups with reverse correlation; panel D displays the results for the groups with salient contingencies. All standard errors are clustered at the participant level.

Figure A1: Treatment effects by decision-making sophistication



Notes: This figure shows the treatment effect on annuity take-up by an index for decision-making sophistication, constructed by standardizing the three comprehension proxies and taking their mean. The treatment effects are relative to the High-price No-Status-Quo group. The text below the arrows reports the difference in treatment effects between participants with above- versus below-median values of the sophistication index. The vertical spikes indicate the 95% confidence interval and standard errors are clustered at the participant level.

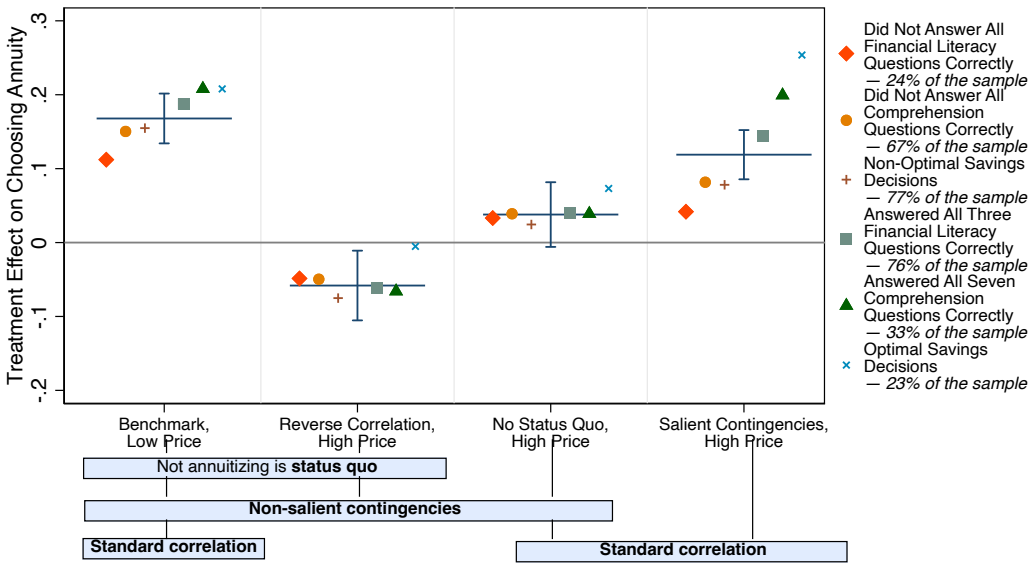
Table A5: Differences in average effects of salient-contingencies treatments on annuity take-up

Treatment Group	Reference Group	Number of Observations	Difference in Effects
Panel A. Across all salient contingencies conditions			
Salient contingencies II, (I + savings specified)	Salient Contingencies I, (savings first)	1,542	–0.035* (0.020)
Salient contingencies III, (II + no context)	Salient Contingencies II, (I + savings specified)	1,523	0.097*** (0.019)
Salient contingencies IV, (III + dominance)	Salient Contingencies III, (II + no context)	1,590	–0.002 (0.017)
Panel B. By savings adjustments in "Salient Contingencies III (II + no context)" and "Salient Contingencies IV (III + dominance)"			
Savings not adjusted — Salient contingencies IV, (III + dominance)	Savings would not be adjusted — Salient Contingencies III, (II + no context)	516	–0.031 (0.026)
Savings adjusted — Salient Contingencies IV, (III + dominance)	Savings would be adjusted — Salient Contingencies III, (II + no context)	1,074	0.014 (0.022)
Savings adjusted, annuity already dominant — Salient Contingencies IV, (III + dominance)	Savings would be adjusted, annuity already dominant — Salient Contingencies III, (II + no context)	729	–0.014 (0.024)
Savings adjusted, annuity not already dominant — Salient Contingencies IV, (III + dominance)	Savings would be adjusted, annuity not already dominant — Salient Contingencies III, (II + no context)	345	0.063 (0.044)

Notes: This table reports the estimates of differences in average treatment effects from a linear probability model of annuity take-up, along with standard errors clustered at the participant level. The difference in treatment effects is estimated as the difference in annuity take-up between the treatment group and the reference group. For participants in the Salient-Contingencies III group who make optimal savings choices, choosing the annuity dominates forgoing the annuity. In Panel B, the groups in rows 4 through 7 are all subsets of Salient Contingencies IV and Salient Contingencies III. For Salient Contingencies IV, the sample descriptions ("savings not adjusted," "savings adjusted, annuity already dominant," "savings adjusted, annuity not already dominant") refer to whether savings was actually adjusted. For Salient Contingencies III, these descriptions refer to how savings would have been adjusted if the adjustment rule in Salient Contingencies IV had been applied to Salient Contingencies III as well. *,**,*** denote differences in treatment effects that are statistically significantly different from 0 at the 10%, 5%, and 1% levels, respectively. All standard errors are clustered at the participant level.

D Heterogeneity Analysis

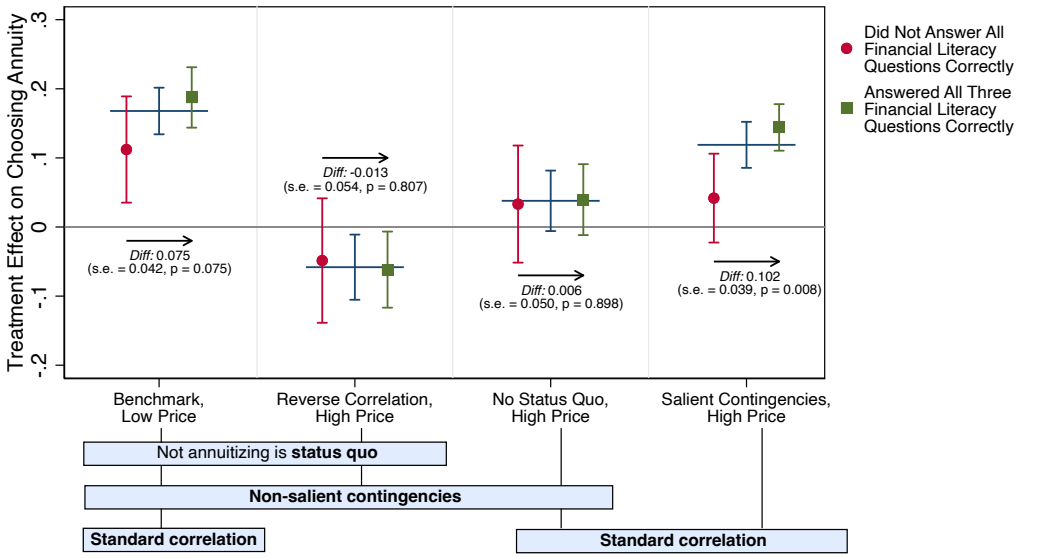
Figure A2: Treatment effects on annuity take-up by financial literacy and comprehension



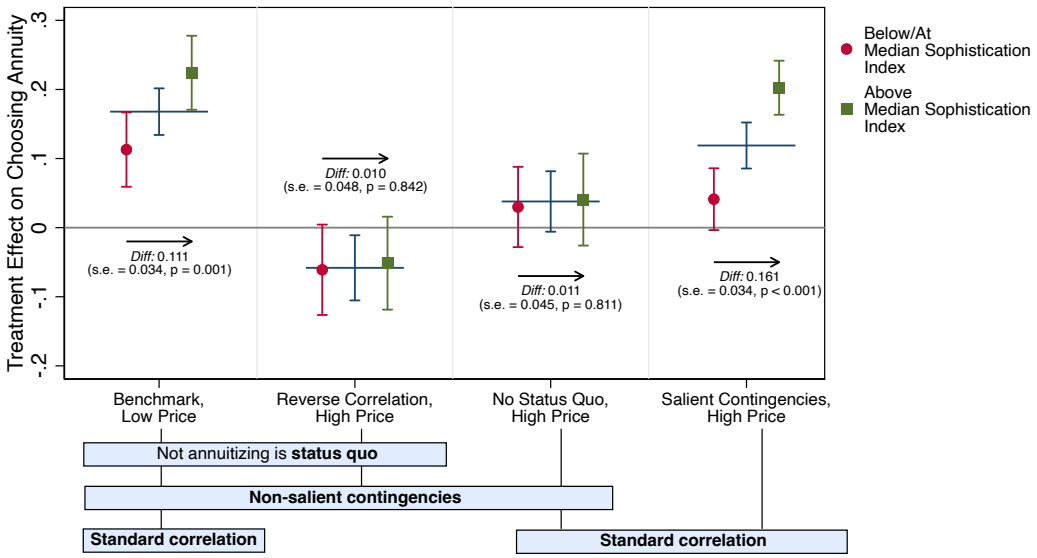
Notes: This figure presents treatment effects, relative to the High-price Benchmark condition, by three proxies for decision-making sophistication.

Figure A3: Heterogeneity of treatment effect on annuity take-up

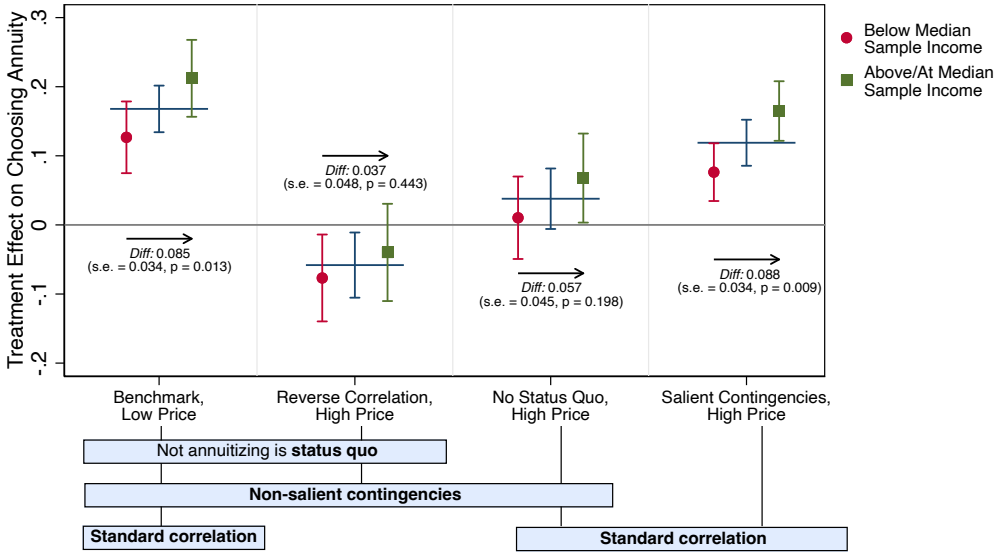
(a) Treatment effect on annuity by accuracy on financial literacy questions



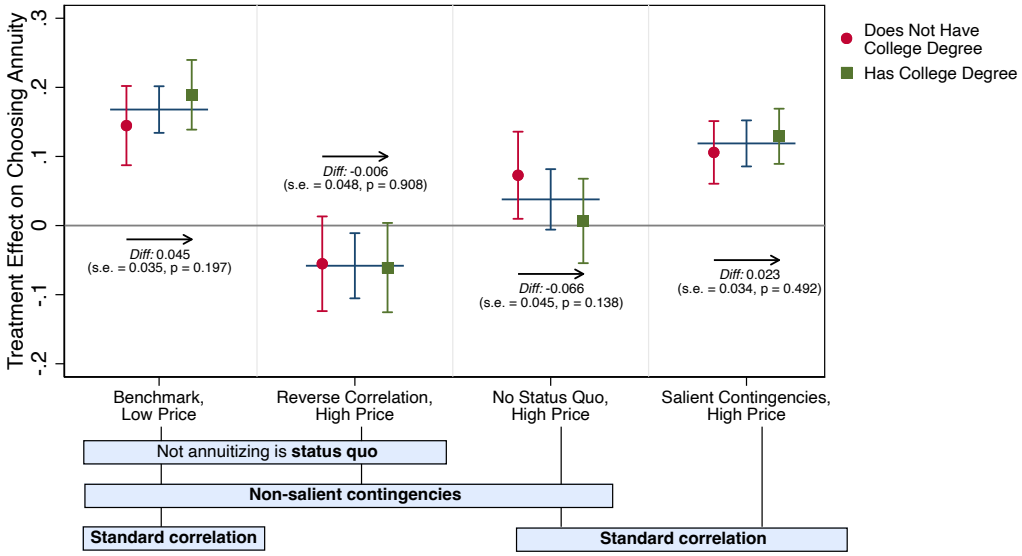
(b) Treatment effects on annuity by comprehension proxy index



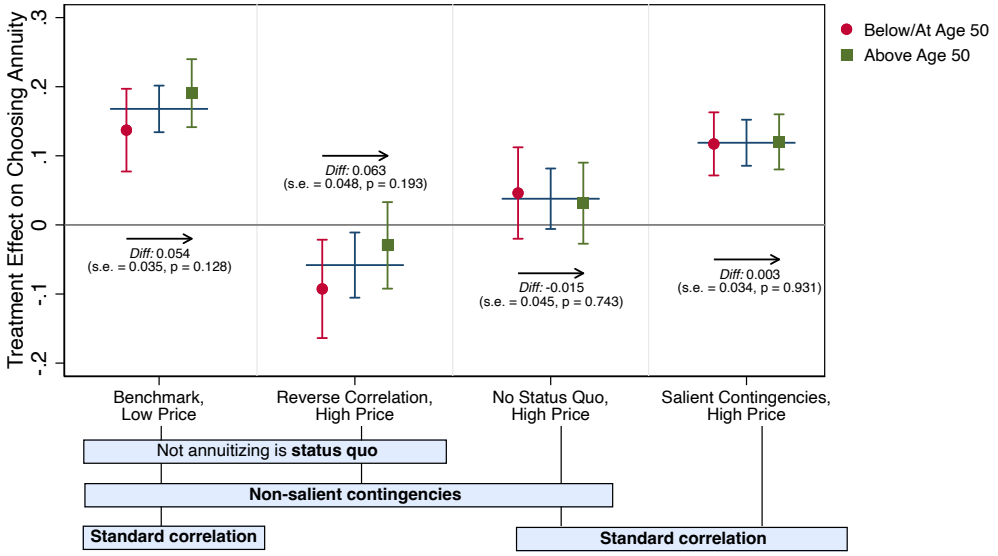
(c) Treatment effect on annuity by income



(d) Treatment effect on annuity by education



(e) Treatment effect on annuity by age



Notes: Panel (a) shows the treatment effects on annuity take-up by whether the participant correctly answered all financial literacy questions in the survey. Panel (b) shows the treatment effects on annuity take-up by a sophistication index, constructed by standardizing the underlying values of the three comprehension proxies (i.e., the fraction of financial literacy questions answered correctly, the fraction of payoff-maximizing savings choices made, and the fraction of comprehension questions answered correctly) and taking their mean. Panel (c) shows the treatment effects on annuity take-up by a binary split across the median income in the sample. Panel (d) shows the treatment effects on annuity take-up by whether the participant has a bachelor's degree. Panel (e) shows the treatment effects on annuity take-up by by a binary split across age 50. The text below the black horizontal bars indicates the difference in treatment effect between the green and red spikes. In all panels, the vertical spikes indicate the 95% confidence interval and standard errors are clustered at the participant level.

Table A6: Interaction effects of treatment and demographic characteristics on annuity take-up

Treatment Group	Reference Group	Number of Participants	Effect of treatment on take-up		
Panel A. Financial Literacy			**Answered all financial literacy questions correctly (N=2,275)**	**Did not answer all financial literacy questions correctly (N=763)**	**Difference**
No Status Quo	Benchmark, high price	1,572	0.040	0.033	0.006
			(0.026)	(0.043)	(0.050)
Salient Contingencies I, (savings first)	No Status Quo	1,530	0.071***	0.016	0.055
			(0.024)	(0.044)	(0.050)
Salient Contingencies II, (I + savings specified)	No Status Quo	1,279	0.055**	–0.065	0.119**
			(0.024)	(0.044)	(0.050)
Salient Contingencies III/IV, (II + no context, dominance)	No Status Quo	1,823	0.145***	0.043	0.101**
			(0.020)	(0.036)	(0.042)
Reverse Correlation	Benchmark, high price	1,508	–0.062**	–0.049	–0.013
			(0.028)	(0.046)	(0.054)
Benchmark, low price	Benchmark, high price	822	0.188***	0.112***	0.075*
			(0.019)	(0.038)	(0.042)
Panel B. Income			**Above/at median income (N=1,444)**	**Below median income (N=1,594)**	**Difference**
No Status Quo	Benchmark, high price	1,572	0.068**	0.010	0.057
			(0.033)	(0.030)	(0.045)
Salient Contingencies I, (savings first)	No Status Quo	1,530	0.078**	0.040	0.037
			(0.030)	(0.029)	(0.042)
Salient Contingencies II, (I + savings specified)	No Status Quo	1,279	0.022	0.025	–0.002
			(0.031)	(0.030)	(0.043)
Salient Contingencies III/IV, (II + no context, dominance)	No Status Quo	1,823	0.142***	0.098***	0.044
			(0.025)	(0.025)	(0.035)
Reverse Correlation	Benchmark, high price	1,508	–0.040	–0.077**	0.037
			(0.036)	(0.032)	(0.048)
Benchmark, low price	Benchmark, high price	822	0.212***	0.127***	0.085**
			(0.025)	(0.024)	(0.034)

Treatment Group	Reference Group	Number of Participants	Effect of treatment on take-up		
Panel C. College Degree			**Has college degree (N=1,623)**	**Does not have college degree (N=1,415)**	**Difference**
No Status Quo	Benchmark, high price	1,572	0.007	0.073**	–0.066
			(0.031)	(0.032)	(0.045)
Salient Contingencies I, (savings first)	No Status Quo	1,530	0.093***	0.018	0.075*
			(0.029)	(0.031)	(0.042)
Salient Contingencies II, (I + savings specified)	No Status Quo	1,279	0.071**	–0.034	0.105**
			(0.029)	(0.032)	(0.043)
Salient Contingencies III/IV, (II + no context, dominance)	No Status Quo	1,823	0.165***	0.069***	0.095***
			(0.024)	(0.025)	(0.035)
Reverse Correlation	Benchmark, high price	1,508	–0.061*	–0.055	–0.006
			(0.033)	(0.035)	(0.048)
Benchmark, low price	Benchmark, high price	822	0.189***	0.145***	0.045
			(0.022)	(0.027)	(0.035)
Panel D. Age			**Above/at age 50 (N=1,773)**	**Below age 50 (N=1,265)**	**Difference**
No Status Quo	Benchmark, high price	1,572	0.033	0.045	–0.012
			(0.029)	(0.034)	(0.045)
Salient Contingencies I, (savings first)	No Status Quo	1,530	0.062**	0.054*	0.008
			(0.028)	(0.032)	(0.043)
Salient Contingencies II, (I + savings specified)	No Status Quo	1,279	0.031	0.012	0.019
			(0.028)	(0.032)	(0.043)
Salient Contingencies III/IV, (II + no context, dominance)	No Status Quo	1,823	0.129***	0.107***	0.021
			(0.023)	(0.027)	(0.036)
Reverse Correlation	Benchmark, high price	1,508	–0.034	–0.090**	0.057
			(0.032)	(0.037)	(0.049)
Benchmark, low price	Benchmark, high price	822	0.187***	0.141***	0.046
			(0.022)	(0.028)	(0.035)

Notes: This table reports estimates of the interaction effects of treatment and various demographic characteristics from a linear probability model of annuity take-up. Panel A displays the results for participants who answered all financial literacy questions correctly and for participants who did not answer all financial literacy questions correctly; panel B displays the results for participants above and below the median income within the sample; panel C displays the results for participants with a college degree and participants without a college degree; panel D displays the results for participants above and below age 50. The treatment effect is estimated as the difference in annuity take-up between the treatment group and the reference group. In row 4 of each panel, participants in the Salient Contingencies III and Salient Contingencies IV groups are pooled. *,**,*** denote estimates that are statistically significantly different from 0 at the 10%, 5%, and 1% levels, respectively. All standard errors are clustered at the participant level.

Table A7: Joint significance of interaction effects of treatments and demographic characteristics on annuity take-up

Treatment Groups	Number of Participants	χ^2 statistic	p-value
Panel A. Financial Literacy			
All Salient Contingencies (I, II, III/IV)	2,352	8.10	0.044
No Status Quo; All Salient Contingencies (I, II, III/IV); Reverse Correlation	3,038	17.42	0.004
Panel B. Income			
All Salient Contingencies (I, II, III/IV)	2,352	2.77	0.428
No Status Quo; All Salient Contingencies (I, II, III/IV); Reverse Correlation	3,038	10.78	0.056
Panel C. College Degree			
All Salient Contingencies (I, II, III/IV)	2,352	8.31	0.040
No Status Quo; All Salient Contingencies (I, II, III/IV); Reverse Correlation	3,038	8.71	0.121
Panel D. Age			
All Salient Contingencies (I, II, III/IV)	2,352	0.46	0.927
No Status Quo; All Salient Contingencies (I, II, III/IV); Reverse Correlation	3,038	2.38	0.794

Notes: This table reports test statistics and p-values from tests of the joint significance of the interaction effects of the treatments listed in the row and the demographic characteristic listed in the panel heading. In all cases, the Salient-Contingencies treatments III and IV are pooled. Hence, the top row of each panel tests for the joint significance of three treatment effects interacted with the listed demographic characteristic and the bottom row of each panel tests for the joint significance of five treatment effects interacted with the listed demographic characteristic. Specifically, panel (a) displays the results for the interaction effect of the listed treatments and of answering all financial literacy questions correctly; panel (b) displays the results for the interaction effect of the listed treatments and of having an income at or above the sample median; panel (c) displays the results for the interaction effect of the listed treatments and of having a college degree; panel (d) displays the results for the interaction effect of the listed treatments and of being above age 50. All standard errors are clustered at the participant level.

E Probit Model Estimation

We estimate the following Probit model:

$$Prob(y_{ij} = 1) = \Phi(\beta_0 + \beta_1 Price_j + \beta_2 NoStatusQuo_j + \beta_3 SalContI_j + \beta_4 SalContIII_j), \quad (1)$$

where $y_{ij} \in \{0, 1\}$ is an indicator that equals 1 if participant i takes up the annuity in condition j, Φ is the cumulative distribution function of the standard normal distribution, $Price_j$ is the price of the annuity in tokens in condition j, $NoStatusQuo_j$ is an indicator variable that equals 1 in the No-Status-Quo and Salient-Contingencies conditions, $SalContI_j$ is an indicator variable for the Salient Contingencies I condition, and $SalContIII_j$ is an indicator variable for the Salient Contingencies III condition. We compute robust standard errors clustered at the participant level. The sample consists of the following experimental conditions: High-price Benchmark, Low-price Benchmark, No-Status-Quo, Salient-Contingencies I, and Salient-Contingencies III.

Appendix Table A8, Column (I) presents the estimates of the Probit model described by equation (1). With these parameters in hand, we rescale the treatment effects to be in units of equivalent price changes in Column (II). Because the error term follows a standard normal distribution, with variance normalized to 1, we obtain the valuations in tokens by dividing the estimated coefficient in Column (I) of the variable indicated in each row by the price coefficient. Column (II) reports these valuations. The row corresponding to the constant term reports the average baseline valuation $\tilde{v}_0$ (in tokens) in the benchmark condition, and the other rows report the change in average valuation (in tokens) caused by the experimental treatment corresponding to the variables in each row. Column (III) reports the average valuations obtained in Column (II) as a share of the average baseline valuation in the benchmark condition ($\tilde{v}_0$). We obtain the estimates in Column (III) by dividing each row in Column (II) by the row corresponding to the constant term in Column (II), which reports $\tilde{v}_0$.

Table A8: Effect of treatments on annuity take-up and valuation

		Average valuation	
	Effect on take-up (I)	In tokens (II)	In % of baseline valuation ($\tilde{v}_0$) (III)
Price (in tokens)	−0.062		
	(0.006)		
No Status Quo	0.115	1.9	6.4%
	(0.068)	(1.0)	(3.7)
Salient Contingencies I	0.198	3.2	11.0%
	(0.072)	(1.2)	(4.0)
Salient Contingencies III	0.458	7.4	25.4%
	(0.076)	(1.4)	(4.5)
Constant	1.805	29.1	
	(0.113)	(1.5)	

Notes: This table reports in column (I) the estimates of average treatment effects from a Probit model of annuity take-up, along with standard errors clustered at the participant level. Price is defined as the annuity price in tokens and is equal to 10 tokens for the low-price annuity and 20 tokens for the high-price annuity. No Status Quo is a binary variable that equals one in the No-Status-Quo condition and in the Salient-Contingencies conditions, and zero in the High- and Low-price Benchmark conditions. The sample consists of the following experimental conditions: High-price Benchmark, Low-price Benchmark, No-Status-Quo, Salient-Contingencies I, and Salient-Contingencies III. Column (II) reports the average baseline valuation $\tilde{v}_0$ (in tokens) in the benchmark condition in the case of the constant term, or the change in average valuation (in tokens) caused by the experimental treatment corresponding to the variables in each row. These measures are obtained by dividing the estimated coefficient of the variable indicated in each row by the price coefficient. Column (III) reports the average valuation obtained in column (II) as a share of the average baseline valuation in the benchmark condition, $\tilde{v}_0$.

F Pre-Analysis Appendix

F.1 Overview

The results in the figures of the body of the paper were all pre-specified in the analysis plan, with two exceptions. First, we added a pooled version of the Salient-Contingencies treatments in Figure 2B because this allowed us to convey the main message of the paper in a single figure without overwhelming the reader with the details of the four different versions of the Salient-Contingencies treatments. Second, we focused the heterogeneity analysis in the body of the paper on the type of heterogeneity for which we had the most interesting results: heterogeneity by decision-making sophistication. The analysis plan pre-specified five dimensions of heterogeneity, including financial literacy and optimality of savings decisions, which are both metrics of decision-making sophistication. We supplemented these two pre-specified metrics with one additional metric of decision-making sophistication, namely answering all comprehension questions correctly. The heterogeneity analysis in the body of the paper shows that the pattern of heterogeneity is the same for each of these three metrics of decision-making sophistication. For the estimation of heterogeneity in treatment effects, we pool the three metrics to increase statistical power.

All pre-specified analyses are reported somewhere in the paper, either in its body or in the appendices. The next two subsections summarize the pre-analysis plan and describe where each of the pre-specified analyses is reported.

F.2 Primary analyses

The pre-analysis plan specified nine main experimental groups. These groups are: 1: Benchmark, High Price (called "G0" in the pre-analysis plan); 2: No Status Quo ("G1"); 3: Salient Contingencies I ("G2"); 4: Salient Contingencies II ("G3"); 5: Salient Contingencies III ("G4"); 6: Salient Contingencies IV ("G5"); 7: Reverse Correlation ("G10"); 8: Benchmark, Low Price ("G20"); and 9: Reverse Correlation, Salient Contingencies IV, Low Price ("G35"). Unless explicitly specified otherwise, the annuity is less than actuarially fair (so "high price") in these nine groups. We pre-specified three sets of primary analyses based on these groups.

First, we specified that we would report the mean annuity take-up in each of these nine main experimental groups, together with the standard errors. This is presented in Figures 2 and 3, and, more comprehensively, in Appendix Table A4.

Second, we pre-specified reporting the average treatment effects of nine comparisons between two experimental groups, G_a, G_b. The nine treatment effects of interest are defined

by the following pairs of reference group (listed first) and treatment group (listed second): 1. G0 vs. G1; 2. G1 vs. G2; 3: G1 vs. G3; 4: G1 vs. G4; 5: G1 vs. G5; 6. G0 vs. G10; 7: G0, "insurance" wording only, vs. G10; 8: G0 vs. G20; and 9: G0 vs. G35. We specified that we would estimate average treatment effects using the following linear probability model:

$$y_{ij} = \beta_0 + \beta_1 \mathbf{1}_{G_b} + \varepsilon_{ij}, \tag{2}$$

where $y_{ij} \in \{0, 1\}$ is an indicator that equals 1 if participant i takes up the annuity in cell $j \in \{G_a, G_b\}$ and where the sample is limited to these two cells. In this and all other regressions, we compute robust standard errors clustered by participant where appropriate (i.e., where some participants appear in multiple cells). Seven of the nine treatment effects listed above are reported in Figures 2 and 3 in the body of the paper. We did not report the remaining two treatment effects in the body of the paper for the following reasons. First, following a pre-specified test from our analysis plan, we designated treatment effect 6 ("G0 vs. G10") as the preferred specification for the Reverse-Correlation effect (and reported it in the body of the paper). The non-preferred specification, treatment effect 7 ("G0, "insurance" wording only, vs. G10") is reported in row 2 of Table A9, but not in the body of the paper. Second, we had included treatment effect 9 ("G0 vs. G35") in the pre-analysis plan to illustrate ceiling effects. This treatment included everything that we expected to increase annuity take-up. Contrary to our expectations, however, the Reverse-Correlation treatment did not increase annuity take-up and, as a result, the largest take-up did not occur in treatment group G35. Hence, this treatment lost it relevance as an illustration of ceiling effects. We therefore reported it in row 2 of Table A9 rather than in the body of the text.

Third, we pre-specified that we would report differences in the average effects of Salient-Contingency treatments on annuity take-up, in the entire sample and for select subsamples. These differences in treatment effects are presented in Appendix Table A5.

Table A9: Average treatment effect on annuity take-up

	Treatment group	Reference group	Effect of treatment on take-up
1	Reverse Correlation	Benchmark, high price	–0.058** (0.024)
2	Reverse Correlation	Benchmark, high price, "insurance" wording only	–0.070** (0.032)
3	Reverse Correlation, low price, Salient Contingencies IV	Benchmark, high price	0.140*** (0.021)
4	Benchmark, high price, "insurance" wording only	Benchmark, high price, other wording	0.019 (0.033)

Notes: This table reports the estimates of average treatment effects from a linear probability model of annuity take-up, along with standard errors clustered at the participant level. The treatment effect is estimated as the difference in annuity take-up between the treatment group and the reference group. In row 4, either "annuity" or "Social Security" wording was used in the reference group. *,**,*** denote treatment effects that are statistically significantly different from 0 at the 10%, 5%, and 1% levels, respectively.

F.3 Secondary analyses

The first set of secondary results comprised replicating the primary analyses on a subsample of participants who made optimal savings decisions in both of the following two cases: (i) no annuity and (ii) high-priced annuity. The mean annuity take-up in each of the nine main experimental groups for this subsample is presented in Figure 4 (Panel A) in body of the paper or in Appendix Table A10 below.

Table A10: Mean annuity take-up - Subsample with optimal savings choices

	Group	Annuity take-up mean
1	Benchmark, high price	0.653 (0.034)
2	No Status Quo	0.727 (0.037)
3	Salient Contingencies I	0.893 (0.024)
4	Salient Contingencies II	0.890 (0.025)
5	Salient Contingencies III and IV	0.922 (0.014)
6	Reverse Correlation	0.648 (0.035)
7	Benchmark, low price	0.861 (0.024)
8	Reverse-Correlation Salient Contingencies IV, low price	0.940 (0.018)

Notes: This table reports the mean annuity take-up rate and the share of participants who made optimal savings choices in the no annuity and high-price annuity cases for different treatment groups, along with standard errors. The sample for estimating take-up means is restricted to participants with optimal savings choices in the no annuity and high-price annuity cases. For participants in the Salient-Contingencies III group with optimal savings choices, choosing the annuity dominates forgoing the annuity, so we pre-specified that this group should be pooled with the Salient-Contingencies IV group.

The corresponding treatment effects in the subsample with optimal savings choices can be found in Appendix Figure A2 or in Appendix Table A11 below. Finally, the differences in the average effects of Salient-Contingency treatments on annuity take-up for this subsample are presented in Appendix Table A12 below.

Table A11: Average treatment effect on annuity take-up - Subsample with optimal savings choices

	Treatment group	Reference group	Effect of treatment on take-up
1	No Status Quo	Benchmark, high price	0.073 (0.050)
2	Salient Contingencies I	No Status Quo	0.167*** (0.044)
3	Salient Contingencies II	No Status quo	0.163*** (0.044)
4	Salient Contingencies III and IV	No Status Quo	0.195*** (0.039)
5	Reverse Correlation	Benchmark, high price	–0.005 (0.049)
6	Reverse Correlation	Benchmark, high price, "insurance" wording only	0.048 (0.069)
7	Benchmark, low price	Benchmark, high price	0.208*** (0.036)
8	Reverse-Correlation Salient Contingencies IV, low price	Benchmark, high price	0.286*** (0.038)
9	Benchmark, high price, "insurance" wording only	Benchmark, high price, other wording	–0.082 (0.072)

Notes: This table reports the estimates of average treatment effects from a linear probability model of annuity take-up, along with standard errors clustered at the participant level. The treatment effect is estimated as the difference in annuity take-up between the treatment group and the reference group. In row 9, either "annuity" or "Social Security" wording was used in the reference group. The sample is restricted to participants with optimal savings choices in the no annuity and high-price annuity cases. For participants in the Salient-Contingencies III group who made optimal savings choices, choosing the annuity dominates forgoing the annuity, so we pre-specified that this group should be pooled with the Salient-Contingencies IV group. *,**,*** denote treatment effects that are statistically significantly different from 0 at the 10%, 5%, and 1% levels, respectively.

Table A12: Differences in average effects of treatment on annuity take-up - Subsample with optimal savings choices

	Treatment group	Reference group	Difference in effects
1	Salient Contingencies II	Salient Contingencies I	–0.004 (0.034)
2	Salient Contingencies III and IV	Salient Contingencies II	0.032 (0.028)

Notes: This table reports the estimates of differences in average treatment effects from a linear probability model of annuity take-up, along with standard errors clustered at the participant level. The difference in treatment effects is estimated as the difference in annuity take-up between the treatment group and the reference group. The sample is restricted to participants who made optimal savings choices in the no annuity and high-price annuity cases. For participants in the Salient-Contingencies III group who made optimal savings choices, choosing the annuity dominates forgoing the annuity, so we pre-specified that this group should be pooled with the Salient-Contingencies IV group. *,**,*** denote differences in treatment effects that are statistically significantly different from 0 at the 10%, 5%, and 1% levels, respectively.

Second, we pre-specified the analysis of the interaction of six selected treatment effects with four dummy variables for: answering all three financial literacy questions correctly, having an income above the median in our sample, having a college degree, and being older than 50 years. To test for the interaction of a particular treatment with a particular demographic covariate d, we run the regression

$$y_{ij} = \beta_0 + \beta_1 \mathbf{1}_{G_b} + \beta_2 d + \beta_3 \mathbf{1}_{G_b} \cdot d + \varepsilon_{ij} \tag{3}$$

with robust standard errors clustered by participant where appropriate. The coefficient β_3 corresponds to the interaction effect of interest. Appendix Table A6 presents the interaction effects of the 4 demographic dummy variables with these these six pre-specified treatment effects.

Because the power of the interaction-effect tests above may be limited, we pre-specified that we would also run tests for interaction effects using specifications that pool several treatment effects. We run four types of tests for each of the four demographic covariates. In the first set of tests, we test for the joint significance of the interaction of one of the four demographic covariates with five pre-specified treatment effects. In the second set of tests, we test for the joint significance of the interaction of one of the four demographic covariates with three Salient-Contingencies treatment effects. These two sets of tests can be found in Appendix Table A7. We also test whether there is heterogeneity with respect to d in which

Salient-Contingencies treatments elicit the strongest response. In the third set of tests we compare the treatment effect of Salient Contingencies III and IV (pooled) relative to No Status Quo and Salient Contingencies I (pooled) and interact this treatment effect with one of the four demographics. In the fourth set of tests, we compare the treatment effect of No Status Quo and Salient Contingencies IV (pooled) relative to Salient Contingencies I and II (pooled) and interact this treatment effect with one of the four demographics. These two sets of tests are presented in Appendix Table A13 below.

Table A13: Heterogeneity of interaction effects of pooled treatments and demographic characteristics in annuity take-up

Pooled Treatment Groups	Pooled Reference Groups	Effect of pooled treatments on take-up		
Panel A. Financial Literacy		**Answered all financial literacy questions correctly**	**Did not answer all financial literacy questions correctly**	**Difference**
Salient Contingencies III and IV	No Status Quo and Salient Contingencies I	0.107***	0.035	0.072**
		(0.015)	(0.029)	(0.033)
Salient Contingencies I and II	No Status Quo and Salient Contingencies IV	–0.016	–0.040	0.025
		(0.016)	(0.030)	(0.034)
Panel B. Income		**Above/at median income**	**Below median income**	**Difference**
Salient Contingencies III and IV	No Status Quo and Salient Contingencies I	0.104***	0.078***	0.026
		(0.019)	(0.019)	(0.027)
Salient Contingencies I and II	No Status Quo and Salient Contingencies IV	–0.028	–0.015	–0.013
		(0.020)	(0.020)	(0.028)
Panel C. College Degree		**Has college degree**	**Does not have college degree**	**Difference**
Salient Contingencies III and IV	No Status Quo and Salient Contingencies I	0.116***	0.061***	0.055**
		(0.018)	(0.020)	(0.027)
Salient Contingencies I and II	No Status Quo and Salient Contingencies IV	–0.009	–0.038*	0.029
		(0.019)	(0.022)	(0.029)
Panel D. Age		**Above/at age 50**	**Below age 50**	**Difference**
Salient Contingencies III and IV	No Status Quo and Salient Contingencies I	0.097***	0.080***	0.016
		(0.018)	(0.021)	(0.027)
Salient Contingencies I and II	No Status Quo and Salient Contingencies IV	–0.023	–0.019	–0.004
		(0.019)	(0.022)	(0.029)

Notes: This table reports estimates of the interaction effects of treatment and financial literacy from a linear probability model of annuity take-up, along with standard errors clustered at the participant level. The treatment effect is estimated as the difference in annuity take-up between the pooled treatment groups and the pooled reference groups. *,**,*** denote estimates that are statistically significantly different from 0 at the 10%, 5%, and 1% levels, respectively.