Proceedings of ICFTBA 2025 Symposium: Data-Driven Decision Making in Business and Economics
DOI: 10.54254/2754-1169/2025.BL30018

When Nudges Backfire: The Effect of Social Norms, Framing Effects, and Default Options on the Pension Saving Decisions in China

Yuming Dong[1*], Zhaoyu Wang[2], Zhuting He[3]

[1]RCF Experimental School, Beijing, China
[2]Beijing Huijia Private School, Beijing, China
[3]Singapore American School, Singapore, Singapore
**Corresponding Author. Email: donyuming@rdfzcygj.cn*

Abstract. This study investigates the impact of behavioral nudges—social norms, loss aversion framing, default options, and positive narrative—on voluntary pension saving decisions in China. Using a randomized controlled trial administered via an online survey with 220 participants, we examine how these interventions influence both the willingness to contribute and the proportion of income allocated to pension savings. Contrary to prevailing international evidence, our results indicate that all nudges significantly reduce willingness to contribute relative to a control group, with the most substantial backfire effects observed for positive narrative and default option nudges. Furthermore, we identify important heterogeneity effects across income groups and emphasize the moderating roles of education, gender, and financial motivation. These findings underscore the critical importance of cultural, institutional, and individual factors in the design and implementation of behavioral interventions. Policy measures tailored to specific demographic segments and focused on financial education and trust-building are recommended.

Keywords: Behavioral Nudge, Pension Savings, Randomized Controlled Trial, China, Household Finance

1. Introduction

The persistent inadequacy of retirement savings represents a critical socioeconomic challenge for aging populations across the globe. According to the OECD, many countries are struggling with pension systems that fail to provide sufficient post-retirement income, leading to increased old-age poverty and fiscal pressure on public budgets. In China, this issue is particularly acute. The country is not only experiencing rapid demographic aging but is also in the midst of substantial pension system reforms. Despite the availability of financial incentives and voluntary contribution schemes, participation rates remain strikingly low [1].

This study investigates how behavioral and socioeconomic factors influence individuals' decisions to contribute to voluntary pension plans. Using original survey data from 220 respondents, we analyze the relationship between various nudges, individual characteristics, and contribution

DOI: 10.54254/2754-1169/2025.BL30018

behavior. Specifically, we examine the effects of social norms, positive and negative framing, and default options—common behavioral interventions—on the willingness to contribute and the contribution amount.

Nudges, as defined by Thaler and Sunstein, are policy tools designed to steer people's decisions in a beneficial direction without restricting freedom of choice. Despite their widespread application in Western contexts, the effectiveness of these interventions in China's unique socio-institutional environment remains underexplored. Our study aims to fill this gap [2].

Contrary to expectations derived from the international literature, our results indicate that all tested nudges significantly reduce willingness to contribute compared to a control group. This backfire effect suggests that widely endorsed behavioral tools may produce unintended consequences in certain contexts. Our findings highlight the importance of cultural and institutional sensitivity in the design and implementation of nudge-based policies and call for a more tailored approach to retirement saving encouragement in China.

The remainder of the paper is structured as follows: Section 2 reviews the relevant literature on nudges and pension savings; Section 3 describes the data and empirical strategy; Section 4 presents the results; and Section 5 discusses the implications and concludes.

2. Literature review

2.1. Theoretical foundations of nudges and behavioral economics

The theoretical underpinnings of nudging emerge from the field of behavioral economics, which integrates psychological insights into economic models to better understand how people make decisions. Herbert Simon's concept of bounded rationality challenged the classical assumption of the perfectly rational homo economicus, suggesting instead that humans operate under cognitive limitations and use simplifying heuristics to navigate complex decisions [3]. This idea was systematically developed by Tversky and Kahneman, who identified a series of cognitive biases—such as the availability heuristic, representativeness, and anchoring—that consistently lead to deviations from rational choice [4]. Their work provides a scientific basis for understanding why people often fail to make the best decisions in the context of uncertainty, time and social influence.

Based on these insights, Thaler and Sunstein introduced the concept of promotion as a practical tool to improve decision-making within the framework they call liberal patriarchy [2]. Promotion is defined as any aspect of the choice structure, which can predictably change people's behavior without prohibiting any choice or significantly changing their economic incentives. This method respects individual freedom and recognizes systemic cognitive deficiencies. Promote the aim of cooperating with human psychology, not opposing human psychology - using prejudices such as inertia, social obedience and loss aversion to guide people to achieve better results in areas from health and finance to environmental protection.

2.2. Empirical evidence on nudges and retirement savings

More and more empirical studies, mainly from Western economies, have proven the powerful impact of promoting retirement savings behavior. The most extensive research and the most effective promotion is the default option. The pioneering research of Madrian and Shea shows that changing the default value of the 401(k) plan in the United States from optional to automatic registration will lead to a significant increase in the participation rate [5]. This influence is mainly driven by the current situation bias - the tendency to insist on preset options due to inertia, lack of

DOI: 10.54254/2754-1169/2025.BL30018

concentration or implicit recognition by default. The success of default rules has been replicated in various settings, most notably in the UK's nationwide auto-enrolment pension policy, which increased pension coverage from 42% to 86%, with the largest gains among low-income and younger workers [6]. However, critics note that defaults can also induce passive decision-making, wherein individuals often remain at suboptimal default contribution rates without engaging in active planning [7].

The framework effect represents another effective drive. According to the prospect theory of Kahneman and Tversky, which determines that individuals attach more importance to losses than equivalent returns, some studies show that the loss framework information is particularly effective in encouraging savings [8]. For example, Gam et al. found that compared with the income framework appeal, the intention to build retirement savings in terms of avoiding future income losses increased by 28% [9]. Similarly, a negative framework that emphasizes the risk of insufficient savings is often better than a positive message that emphasizes income.

Social norms - providing information about the behavior of peers - are also widely used in the field of savings. Duflo and Saez show that the participation rate of informing employees and colleagues in retirement plans has significantly increased the registration rate [10]. Dur et al. conducted a field experiment in a retail bank, and customers who received information on the average savings balance of their peers then increased their savings [11]. These interventions use the power of descriptive social norms to reduce pluralistic ignorance and signal socially recognized behavior.

2.3. Mechanisms of failure: when and why nudges backfire

Although they have proven effective in many cases, pushing does not always produce the expected effect, and sometimes leads to negative results. Psychological response is the key mechanism to promote possible failures. Rooted in the theory of Brehm, when individuals think that confersions threaten their freedom of choice, they will react, causing them to deliberately act in the opposite way to the recommended actions [12,13]. This is especially possible when the thrust is too obvious or considered manipulative.

The thrust may also be counterproductive, causing discouragem or loss of motivation. Beshears et al. pointed out that when individuals are pushed to a seemingly unattainable goal - such as saving amounts far beyond their ability - they may be completely disconnected [7]. Bhargava and Manoli provide evidence that social comparison is particularly frustrating for low-income individuals, who may interpret the high savings rates of others as evidence that they cannot save, leading to reduced efforts [14].

In addition, social image problems may cause unexpected reactions. Bolton et al. developed a theoretical model that shows that if contribution is considered boasting or prohibition is considered socially acceptable, making behavior observable may be counterproductive. In this case, relying on the promotion of social popularity may reduce the participation of individuals who are sensitive to images [15].

Several empirical studies have recorded this counterproductive effect. Sunstein and Thunström et al. reviewed the cases that push not only failed to change the behavior, but also worsened the results, and emphasized the importance of background factors, individual differences and cultural norms in mediating the effect of the recommendation [16,17].

DOI: 10.54254/2754-1169/2025.BL30018

2.4. The Chinese context: a distinctive setting for nudge effectiveness

China's research on behavior promotion, especially on pension savings, is still limited. China's pension system is highly decentralized, combining basic public pensions with voluntary corporate and individual plans. In addition, financial literacy is relatively low, and families show high savings rates mainly out of preventive motives, not retirement plans [18]. Cultural factors such as collectivism, familial responsibility, and trust in authorities may also shape how nudges are received. For example, social norms may exert stronger influence in China's collectivist culture, but could also provoke stronger reactance if perceived as invasive. Default options might be more effective in a high-trust environment, but could also be ignored if institutions are distrusted.

Existing studies on Chinese pension behavior have focused mainly on structural and economic determinants, with scant attention to behavioral mechanisms or the potential of nudges. This gap is significant given the urgent need to enhance retirement preparedness in a rapidly aging society. By examining the effects of social norms, framing, and defaults in a controlled setting while accounting for demographic and socioeconomic moderators, this study aims to provide much-needed evidence on the applicability and limitations of nudge theory in the Chinese context.

The aforementioned international evidence establishes the efficacy of nudges under certain conditions, primarily in Western, high-trust institutional settings. However, the unique socio-institutional fabric of China suggests that these interventions may not translate directly and could even produce counterintuitive results. The potential for psychological reactance is heightened in a context where trust in financial institutions is still evolving, and individuals may perceive nudges from official sources as coercive rather than helpful. Furthermore, social norm messages might not only invoke conformity but also trigger feelings of inadequacy or resentment in a highly competitive and unequal society. Similarly, default options, which rely on institutional trust and inertia, may be met with suspicion if the underlying system is perceived as unreliable or unfair. Consequently, while we test the standard hypotheses derived from the international literature, we must also seriously entertain the possibility of null or even backfiring effects, where nudges decrease willingness to participate.

2.5. Hypothesis

Based on the extensive international evidence, we derive the following primary hypotheses:

- H1a: Social norm nudges increase willingness to contribute to pensions.
- H1b: Loss-framing nudges increase willingness to contribute.
- H1c: Default-option nudges increase willingness to contribute.
- H1d: Positive-narrative nudges increase willingness to contribute.

However, the discussion of China's unique context (Section 2.4)—characterized by potential trust deficits and different social dynamics—suggests that these hypothesized positive effects may be attenuated or even reversed.

2.6. Conceptual framework and research contributions

Guided by the literature, we propose a conceptual framework in which behavioral nudges influence saving intentions through cognitive and social pathways—including social comparison, loss aversion, and inertia—while also considering moderating factors such as income, education, financial literacy, and trust. Unlike most previous studies, we explicitly allow for the possibility of null or negative effects, acknowledging that nudges are not universally beneficial.

DOI: 10.54254/2754-1169/2025.BL30018

This study contributes to the literature in several ways: it tests established behavioral interventions in an understudied cultural context; it examines both average effects and heterogeneous responses across subgroups; and it explores why nudges may fail, thereby addressing a critical gap in both the theoretical and practical understanding of behavioral public policy.

3. Methodology

3.1. Research design

This study employs a randomized controlled trial (RCT) design embedded within a cross-sectional survey to identify the causal effects of behavioral nudges on pension saving intentions. Respondents were randomly assigned to one of the four treatment groups. Each treatment group received a distinct nudge intervention: (1) Social Norms, (2) Loss Aversion Framing, (3) Default Option, or (4) Positive Narrative Framing. The control group received neutral, reference-only baseline information about the voluntary pension plan, without any behavioral factors. This design allows a clear comparison between the effectiveness and baseline of each push, and randomization ensures that the difference in the observed results can be attributed to the intervention, not the original characteristics of the respondents.

3.2. Survey instrument and nudge implementation

The questionnaire is divided into four main parts: (1) informed consent and introduction, (2) socio-economic and population overview, (3) random prompt intervention, and (4) result measures and follow-up questions.

The promotion of interventions is carefully designed in accordance with the established protocols in the behavioral economics literature, and pre-tested with pilot samples to ensure clarity and cultural appropriateness.

(1) Social Norm Tips: Show the following statement to the participants: Your friends and colleagues are replenishing your pension fund. This uses descriptive social norms to signal the behavior of peers.

(2) Loss aversion tips: This intervention adopts a loss framework message: if you don't replenish your pension fund, your quality of life will decline in the future. This framework highlights potential losses in the future.

(3) Default option tips: Participants are informed that if you choose to participate, you will automatically register for the default payment of the pension savings plan.You can freely adjust or cancel this setting at any time. This leverages status quo bias and inertia.

(4) Positive Narrative Nudge: This group received a gain-framed message featuring a short, relatable story: If you supplement your retirement funds, your retirement life will be greatly improved, and you will have more freedom of choice. This uses a positive role model and outcome imagery.

The control group received a neutral statement: " A voluntary pension plan is available for you to consider. Please decide whether to participate based on your situation."

3.3. Data collection and sampling

Data were collected via an online survey platform between July 30, 2025 and August 2, 2025. The sample was recruited to roughly mirror the urban Chinese adult population (aged 25-60) with internet access, using gender, age, and income quotas based on national statistics to enhance

DOI: 10.54254/2754-1169/2025.BL30018

representativeness. The final sample consisted of 220 respondents. After random assignment, each treatment group contained approximately 42 individuals, and the control group contained 52 individuals.

To ensure data quality, attention checks were embedded within the survey (e.g., "Please select 'Strongly Agree' for this question"). Responses that failed these checks or were completed in an unrealistically short time were excluded from the final analysis, resulting in the analytical sample of N=210.

3.4. Variable measurement

Dependent Variables:

Willingness to Contribute: Measured on an 11-point Likert scale (0 = 'Extremely Unwilling' to 10 = 'Extremely Willing') in response to the question: "How willing are you to participate in a voluntary pension plan?"

Contribution Proportion: An ordinal variable measured on an 11-point scale, elicited by the question: "If you were to participate, what percentage of your monthly income would you be willing to contribute?" Respondents selected a number on the 0-10 scale, with each point representing a decile of income.

Independent Variables:

The primary independent variables are binary indicators for assignment to each of the four nudge conditions.

Control Variables:

A comprehensive set of covariates was collected to account for potential confounding factors and to test for heterogeneity:

(1) Socio-demographics: Age, gender, and education level.

(2) Economic Factors: Monthly income, employment status.

(3) Geographic Region: Coded into administrative regions for control.

3.5. Empirical strategy

The analysis is carried out in two main stages.

First of all, we estimated the following ordinary least squares (OLS) regression model to evaluate the average treatment intention effect of the promotion:

$$Y_i = \beta_2 * X + \gamma * C_i + \varepsilon_i \quad (1)$$

Where Y_i is the outcome variable (Willingness or Proportion) for individual i, the $\beta_2$captures the effect of each nudge relative to the control group; X is independent variables, including (1) Social Norms, (2) Loss Aversion Framing, (3) Default Option, or (4) Positive Narrative Framing. C_i is the vector of control variables, and ε_i is the error term.

4. Results

4.1. Descriptive statistics and balance check

The analysis first checked the summary statistics and checked the balance between the treatment group and the control group. As shown in Table 1, the final analysis sample consists of 210

DOI: 10.54254/2754-1169/2025.BL30018

respondents. The sample is divided equally by gender, and the average age is between 35 and 44 years old. The average education level is slightly higher than that of an associate's degree, and the average monthly income range corresponds to the category of 5,000-10,000 yuan. Crucially, the randomization procedure was successful. As evidenced by the closely aligned means between the overall sample and the treatment group for all key demographic and socioeconomic variables, there were no systematic differences between the groups before the administration of the nudge interventions. This ensures that any subsequent differences in outcomes can be confidently attributed to the treatments rather than to pre-existing sample characteristics.

Table 1. Summary statistics

Variable	Obs	Mean	Std. dev.	Min	Max
Overall sample					
edu	210	2.238	1.272	1	4
region	210	8.667	3.874	1	17
age_bands	210	3.476	1.261	1	5
female	210	0.500	0.501	0	1
income_bands	210	4.119	1.871	1	6
employ	210	1.357	0.686	1	4
Incentives motivation	216	0.778	0.417	0	1
Loss Aversion	210	0.200	0.401	0	1
Social Norm	210	0.200	0.401	0	1
Default Option	210	0.200	0.401	0	1
Positive Narrative	210	0.200	0.401	0	1
Treatment group					
edu	42	2.238	1.284	1	4
region	42	8.667	3.912	1	17
age_bands	42	3.476	1.273	1	5
female	42	0.500	0.506	0	1
income_bands	42	4.119	1.890	1	6
employ	42	1.357	0.692	1	4

Notes: This table reports summary statistics for the full analytical sample (N=210) and the pooled treatment group. The variables include: education level (edu, 1=High school or below to 4=Master's or above), region (region, 1-17), age bands (age_bands, 1=25-34 to 5=55-60), gender (female, 1=female, 0=male), monthly income bands (income_bands, 1=<¥5,000 to 6=>¥25,000), employment status (employ, 1=Full-time to 4=Unemployed/Other), and binary indicators for assignment to each nudge condition. The close alignment of means between the full sample and the treatment group across all observable characteristics confirms that the randomization procedure was successful and the groups are well-balanced at baseline.

4.2. Main effects of nudges on willingness to contribute

Model 1 estimates the intent-to-treat effects without control variables. The coefficient for the composite Baseline_intent (representing the control group's baseline) is positive, but not the focus. The key findings are the coefficients for the nudge treatments. All four nudges show negative coefficients compared to the control group. The positive narrative nudge exhibits the largest negative effect, followed by the default option. The social norm and loss aversion nudges also

DOI: 10.54254/2754-1169/2025.BL30018

display negative directional effects but are not statistically significant at conventional levels in this specification.

The stability of the coefficients for the nudges of Model 1 confirms the robustness of these findings and suggests that the observed backfire effect is not driven by omitted variable bias.

Control Variables: The covariates reveal important patterns. Being female is associated with a significantly higher willingness to contribute. Conversely, higher income, being employed, and older age are all associated with a statistically significant decrease in willingness. Education and region did not show a significant relationship with willingness in this model.

Table 2. Willingness (treatment groups)

Dependent variable: Willingness				
	(1)	(2)	(3)	(4)
Control Group	-0.970*	-0.970**		
	(0.498)	(0.433)		
Loss Aversion			-0.881	-0.881
			(0.634)	(0.552)
Social Norm			-0.833	-0.833
			(0.634)	(0.552)
Default Option			-1.024	-1.024*
			(0.634)	(0.552)
Positive Narrative			-1.143*	-1.143**
			(0.634)	(0.552)
edu		0.138		0.138
		(0.146)		(0.147)
region		0.028		0.028
		(0.048)		(0.048)
age_bands		-0.356**		-0.356**
		(0.146)		(0.147)
female		1.275***		1.275***
		(0.353)		(0.355)
income_bands		-0.472***		-0.472***
		(0.103)		(0.103)
employ		-0.899***		-0.899***
		(0.287)		(0.289)
Constant	6.286***	9.499***	6.286***	9.499***
	(0.445)	(0.976)	(0.448)	(0.982)
N	210	210	210	210
R-sq	0.018	0.278	0.019	0.279

Notes: This table presents OLS regression estimates for the effect of behavioral nudges on willingness to contribute (scale 0-10). Models 1 and 3 include only the nudge treatment dummies. Models 2 and 4 add controls for education, region, age, gender, income, and employment status. The omitted category is the control group. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

DOI: 10.54254/2754-1169/2025.BL30018

4.3. Effects on contribution proportion

Table 3 reports the regression results for the second outcome variable: the proportion of monthly income respondents were willing to contribute. In stark contrast to the findings on willingness, none of the nudge treatments show a statistically significant effect on the contribution proportion in Model 1. All coefficients are positive but very small in magnitude and are statistically indistinguishable from zero. This indicates that while the nudges negatively impacted the general willingness to participate, they did not influence the intended amount of contribution among those who were willing.

Table 3. Proportion

Dependent variable: Proportion				
	(1)	(2)	(3)	(4)
Control Group	0.113	0.113		
	(0.485)	(0.444)		
Loss Aversion			0.048	0.048
			(0.618)	(0.566)
Social Norm			0.095	0.095
			(0.618)	(0.566)
Default Option			0.119	0.119
			(0.618)	(0.566)
Positive Narrative			0.190	0.190
			(0.618)	(0.566)
edu		0.456***		0.456***
		(0.150)		(0.151)
region		-0.076		-0.076
		(0.049)		(0.050)
age_bands		-0.085		-0.085
		(0.149)		(0.150)
female		0.435		0.435
		(0.362)		(0.364)
income_bands		-0.537***		-0.537***
		(0.105)		(0.106)
employ		-0.150		-0.150
		(0.294)		(0.296)
Constant	4.857***	6.994***	4.857***	6.994***
	(0.434)	(1.000)	(0.437)	(1.008)
N	210	210	210	210
R-sq	0.000	0.186	0.001	0.186

Notes: This table presents OLS regression estimates for the effect of behavioral nudges on the proportion of monthly income respondents are willing to contribute. Models 1 and 3 include only the nudge treatment dummies. Models 2 and 4 add controls for

DOI: 10.54254/2754-1169/2025.BL30018

education, region, age, gender, income, and employment status. The omitted category is the control group. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

The control variables again provide meaningful insight. Education is a powerful positive predictor, suggesting that each higher level of education is associated with an increase of 0.456 points on the 10-point scale measuring contribution proportion. Income, however, remains a strong negative predictor, reinforcing the finding that higher-income groups, while potentially more capable of saving, exhibit greater resistance to increasing their pension contribution rates. Age, gender, and employment status were not significant predictors of the contribution proportion.

4.4. Heterogeneity analysis: high-income vs. low-income groups

Given the strong negative relationship between income and our outcome variables, we conducted a subgroup analysis by splitting the sample into high-income and low-income groups based on the median income band. The results, presented in Table 4, reveal notable differential effects. For willingness to contribute, the negative impact of the nudges appears more pronounced and statistically significant among the high-income subgroup. The coefficient for the positive narrative nudge, for instance, is significant at the 10% level for both groups but larger for the high-income group.

Table 4. Willingness (income groups)

Dependent variable: Willingness				
	(1)	(2)	(3)	(4)
	High income	Low income	High income	Low income
Control Group	-1.115**	-1.048*		
	(0.538)	(0.590)		
Loss Aversion			-1.154	-0.846
			(0.698)	(0.754)
Social Norm			-1.077	-0.885
			(0.698)	(0.754)
Default Option			-1.000	-1.192
			(0.698)	(0.754)
Positive Narrative			-1.231*	-1.269*
			(0.698)	(0.754)
Education	-0.132	0.337*	-0.132	0.337*
	(0.255)	(0.199)	(0.261)	(0.201)
Region	-0.294***	0.125**	-0.294***	0.125**
	(0.072)	(0.062)	(0.074)	(0.063)
Age bands	-0.894***	-0.303	-0.894***	-0.303
	(0.285)	(0.191)	(0.293)	(0.193)
female	2.900***	1.071**	2.900***	1.071**
	(0.519)	(0.506)	(0.532)	(0.512)
Income bands	0.600**	-1.234**	0.600*	-1.234**
	(0.294)	(0.615)	(0.302)	(0.622)

DOI: 10.54254/2754-1169/2025.BL30018

Table 4. (continued)

employ	0.000	-0.912*	0.000	-0.912*
	(.)	(0.471)	(.)	(0.476)
Constant	11.235***	12.245***	11.235***	12.245***
	(1.181)	(3.296)	(1.212)	(3.330)
N	65	130	65	130
R-sq	0.436	0.194	0.437	0.197

Notes: This table reports results from OLS regressions examining heterogeneous treatment effects by income level. The sample is split into high-income (above median) and low-income (below median) groups. All specifications include the full set of control variables (education, region, age, gender, income bands, employment status). Columns (1) & (2) use a binary indicator for any nudge treatment (Incentive motivation). Columns (3) & (4) show estimates for each nudge type. Robust standard errors are in parentheses. on willingness to contribute (measured on a 0-10 point scale).*** p<0.01, ** p<0.05, * p<0.1.

The effects of the control variables also differ substantially by income group: Education has a positive and significant effect on willingness for low-income individuals but a negative, insignificant effect for high-income individuals. Region exhibits contrasting effects: it is negatively associated with willingness in the high-income group but positively associated in the low-income group. Age is a strong negative predictor for the high-income group, but is insignificant for the low-income group. The positive effect of being female is much larger in the high-income group than in the low-income group. For the contribution proportion, the null results for the nudges held consistently across both income subgroups, mirroring the findings from the full sample.

4.5. Summary of key findings

Backfire Effect on Willingness: Contrary to the hypotheses derived from the literature, all tested behavioral nudges exerted a negative influence on the willingness to contribute to a voluntary pension plan. This effect was statistically significant for the positive narrative and default option nudges. No Effect on Contribution Amount: The nudges did not have a statistically significant effect on the proportion of income individuals were willing to contribute, suggesting the backfire effect was on the participation decision rather than the savings intensity. Importance of Socioeconomic Moderators: The results highlight the critical role of individual characteristics. Higher education increased the planned contribution amount, while higher income strongly decreased both willingness and contribution proportion. The effects of nudges and other controls were highly heterogeneous across income groups. Robustness: The estimated effects of the nudges were stable across model specifications with and without controls, lending credibility to the findings.

These results paint a complex picture, indicating that the application of behavioral nudges in the Chinese pension context is not straightforward and can produce unintended consequences. The following discussion section will explore the potential explanations and implications of these findings.

5. Discussion and conclusion

5.1. Interpretation of key findings

This study set out to investigate the effects of behavioral nudges—social norms, loss aversion framing, default options, and positive narrative—on voluntary pension saving decisions in China. Contrary to the core principles of a large number of international literature and promoting theories,

DOI: 10.54254/2754-1169/2025.BL30018

our results have always shown that these interventions reduce the willingness of individuals to contribute. This counterproductive effect is most obvious in the promotion of positive narrative and default options.

Several mechanisms can explain these counterintitive results. First of all, against the background that trust in financial institutions and pension systems is still developing, external attempts to guide behavior may trigger psychological reactions. Individuals may think that these promotions are manipulative or violate their autonomy, causing them to defend their freedom by rejecting suggestions. Secondly, the content promoted may inadvertently cause anxiety or frustration. For example, a loss-aversion framework that emphasizes future income exhaustion may be seen as overwhelming or threatening, especially among those with limited financial capacity, leading to defensive avoidance rather than participation. Similarly, social norms indicate that high savings among peers may create a sense of inadequacy or resentment rather than a positive incentive benchmark.

The stark contrast between our findings and those from Western studies underscores the critical role of cultural and institutional context. China's collectivist social norms, unique pension system structure, and rapidly changing economic environment shape how individuals process information and make financial decisions. A nudge that is effective in an individualistic, high-trust society may not translate directly to a context with different social dynamics and levels of institutional confidence. Furthermore, our heterogeneity analysis revealed that the negative effects were not uniform across the population. The resistance was particularly strong among high-income individuals, who also demonstrated a significantly lower baseline willingness to contribute. This suggests that for those with more resources and potentially more financial sophistication, simplified nudges may be viewed as patronizing or irrelevant. Conversely, the positive effect of education on contribution percentage (an increase of 0.456 points on the 0-10 scale) highlights that cognitive resources and financial literacy are pivotal in facilitating proactive saving decisions, a factor that simple nudges cannot substitute.

5.2. Theoretical and policy implications

Theoretically, our findings contribute to the growing literature on the boundary conditions and potential pitfalls of nudge theory. They serve as a compelling reminder that nudges are not a universal remedy; their effectiveness is deeply contingent on the socio-cultural environment and the specific characteristics of the target population. Researchers must move beyond simply testing if nudges work, and instead investigate when, why, and for whom they work—or fail.

From a policy perspective, our results carry significant implications for the design of retirement savings programs in China and similar contexts: Prioritize Education over Manipulation: Given that education was the strongest positive predictor of saving behavior, policymakers should invest in long-term financial education and literacy programs. Equipping individuals with the knowledge and tools to understand complex financial products is a more sustainable and ethically transparent approach than leveraging cognitive biases.

Abandon One-Size-Fits-All Nudges; Embrace Personalization: The failure of generic nudges and the strong heterogeneity effects suggest that effective interventions must be tailored to specific demographic segments. For example, communication strategies for high-income, financially literate individuals should focus on detailed information, tax benefits, and advanced planning tools, avoiding simplistic prompts. For lower-income groups, messages that reduce complexity, build trust, and highlight attainable goals may be more effective. Proceeding with Caution: Mandatory Policies May Be More Appropriate than Nudges: The consistent backfire effect of the default option nudge—

DOI: 10.54254/2754-1169/2025.BL30018

one of the most powerful tools in other countries—is particularly telling. In an environment where automatic enrollment may be met with suspicion, policymakers might consider soft mandatory policies, such as facilitated enrollment (where individuals must make an active choice) or matching contribution schemes that provide a tangible incentive rather than relying on inertia. Build Trust Before Nudging: The effectiveness of any behavioral intervention is predicated on a foundation of trust. Therefore, policy efforts should first focus on enhancing the transparency, reliability, and perceived fairness of the pension system. Without trust, any attempt to "nudge" will likely be met with skepticism and resistance.

5.3. Limitations and future research

This study has several limitations. First, it measured behavioral intentions rather than actual saving behavior, which may differ. Second, the online sample, while representative of urban internet users, may not generalize to rural populations. Third, the nudges were delivered as one-time messages; their effects might differ in a real-world setting with repeated exposure.

Future research should aim to replicate these findings with behavioral outcomes in field experiments. It should also explore a wider range of potential nudges, such as those that focus on planning prompts, commitment devices, or personalized projections. Investigating the role of social influencers (e.g., family, community leaders) rather than anonymous peers could also be a fruitful avenue, given the importance of familial ties in Chinese culture.

5.4. Conclusion

In conclusion, this study provides robust evidence that commonly endorsed behavioral nudges can have paradoxical effects in the Chinese pension context, reducing the willingness to save for retirement. This underscores the necessity of developing context-sensitive behavioral policies that are grounded in a deep understanding of local institutions, culture, and individual differences. Rather than relying on imported behavioral solutions, policymakers should focus on building financial capability, fostering trust, and designing programs that respect the autonomy and diversity of the population they aim to serve.

6. Ethical considerations and limitations

All participants provided informed consent at the beginning of the survey and were debriefed about the study's purpose after completion.

Several limitations are acknowledged. First, the study measures behavioral intentions rather than actual saving behavior, which may differ. Second, the online sample, while quota-controlled, may not be fully representative of the entire Chinese population, particularly rural residents and those without internet access. Third, the nudges were delivered as one-time messages, whereas their effectiveness might change with repeated exposure or in a real-world choice environment. Finally, while we control for key covariates, the possibility of unobserved confounding factors cannot be entirely ruled out.

References

[1] OECD. (2023). Pensions at a glance 2023: OECD and G20 indicators. OECD Publishing. https://doi.org/10.1787/678055dd-en

DOI: 10.54254/2754-1169/2025.BL30018

[2] Thaler, R. H., & Sunstein, C. R. (2009). Nudge: Improving decisions about health, wealth, and happiness. Penguin Books.
[3] Simon, H. A. (1957). Models of man: Social and rational. Wiley.
[4] Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. Science, 185(4157), 1124–1131.
[5] Madrian, B. C., & Shea, D. F. (2001). The power of suggestion: Inertia in 401(k) participation and savings behavior. The Quarterly Journal of Economics, 116(4), 1149–1187.
[6] Cribb, J., & Emmerson, C. (2022). Automatic enrolment and pension saving in the UK. Economica, 89(355), 2053–2078.
[7] Beshears, J., Choi, J. J., Laibson, D., & Madrian, B. C. (2008). The importance of default options for retirement saving outcomes: Evidence from the United States. In S. J. Kay & T. Sinha (Eds.), Lessons from pension reform in the Americas (pp. 59-87). Oxford University Press.
[8] Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. Econometrica, 47(2), 263–292.
[9] Gam, J., Huang, K., & Zhao, M. (2021). Loss aversion and retirement savings. Journal of Consumer Research, 48(3), 4237–4255.
[10] Duflo, E., & Saez, E. (2003). The role of information and social interactions in retirement plan decisions: Evidence from a randomized experiment. The Quarterly Journal of Economics, 118(3), 815–842.
[11] Dur, R., Fleming, D., van Garderen, M., & van Lent, M. (2021). A social norm nudge to save more: A field experiment at a retail bank. Journal of Public Economics, 200, 104443.
[12] Brehm, J. W. (1966). A theory of psychological reactance. Academic Press.
[13] Osman, M. (2020). Psychological reactance and behavioral economics. In M. Altman (Ed.), Behavioral economics for dummies (pp. 215-230). For Dummies.
[14] Bhargava, S., & Manoli, D. (2015). Psychological frictions and the incomplete take-up of social benefits: Evidence from an IRS field experiment. American Economic Review, 105(11), 3489–3529.
[15] Bolton, P., Brunnermeier, M. K., & Veldkamp, L. (2020). Leadership, coordination, and corporate culture. The Review of Economic Studies, 87(4), 1735–1776.
[16] Sunstein, C. R. (2017). Nudges that fail. Behavioural Public Policy, 1(1), 4–25.
[17] Thunström, L., Nordström, J., Shogren, J. F., Ehmke, M., & van 't Veld, K. (2018). The effect of social norms on residential electricity consumption. Energy Policy, 115, 99-106.
[18] Chen, Y., Liu, T., & Zhang, L. (2021). Pension reform and behavioral incentives in China. Journal of Aging Studies, 55, 100876.