CAMBRIDGE
UNIVERSITY PRESS

ARTICLE

More than a wink and a nudge: examining the choice architecture of online government budget simulations

Whitney Afonso[1] (ID) and Zachary Mohr[2]

[1]School of Government, University of North Carolina, Chapel Hill, NC, USA and [2]School of Public Affairs and Administration, University of Kansas, Lawrence, Kansas
Corresponding author: Zachary Mohr; Email: zmohr@ku.edu

Abstract
Following the growing interest in using behavioral theory and choice architecture in the public sector, several new studies have looked at how changes in the choice architecture of budget simulations influence the participants' budgetary decisions. These studies have also introduced the possible problem that participants may make inappropriate choices in the budget simulation, like creating a budget with unacceptably high budget surpluses. Building on Thaler and Sunstein's NUDGES framework, we seek to answer the question, 'How can budgetary choice architects correct for errors such as large ending surpluses at the end of the budget simulation?' We replicate earlier results on budget starting conditions. Additionally, we test a budget treatment that encourages participants to reduce ending budget surpluses. The budget treatment works as intended and suggests that the large ending budget surpluses stem from errors made by participants in the simulation rather than loss aversion. The need to both nudge and budge participants is important for practicing choice architects, like public budgeters who have to design and implement tools that inform citizens and reveal accurate preferences that conform with legal requirements.

Keywords: budget simulation; choice architecture; citizen engagement; local government

Introduction

Increasingly, behavioral economics and behavioral public policy is becoming more important in the public sector. Professional organizations in the public sector recognize the importance of behavioral theory and have called upon their members to consider how they are choice architects[1] that develop the processes upon which people make decisions (Riis and Peterson, 2022; Kavanagh and Argarunova, 2022). For example,

---

[1]According to Thaler and Sunstein (2021, p. 19), a choice architect is the person who 'has the responsibility for organizing the context in which people make decisions.'

the Government Finance Officers Association (GFOA) notes that decision-makers are human and thus can make choices that are 'inconsistent and biased,' and call upon their members to design processes that recognize decision heuristics like status quo bias and the importance of defaults (Riis and Peterson, 2022). Likewise, scholars who study public management (Linos *et al.*, 2020) and policy (Castleman *et al.*, 2021) have also increasingly been drawn to the literature of choice architecture and nudging, but the literature on behavioral public budgeting and financial management is much more limited and has not kept up with the demands of practice (Mohr and Kearney, 2021).

Early tests of behavioral theory in public budgeting environments, like budget simulations, show promise. Tuxhorn *et al.* (2022) showed that dynamic online simulations that provide feedback to participants produce different budgetary preferences than static simulations. Mohr and Afonso (2024) show that the default starting conditions also change the participants' budgetary preferences, but this research also shows that under certain conditions, people are likely to develop unacceptably high surpluses[2] at the end of their budget simulations. In this research, we develop a treatment to address how to reduce the end of budget simulation surpluses while maintaining users' autonomy and freedom to make a reasoned budgetary choice.[3]

We study this question in the context of a Balancing Act online budget simulation, which is used extensively in local governments to engage citizens with the budget process. The results show that changing the initial starting conditions influences budget choices in many ways that are similar to previous findings. We also show that the treatment to reduce the final budgetary surplus worked as intended. This supports the NUDGES premise that simulation user error is more likely the causal mechanism than loss aversion. In the 'Discussion' section, we discuss other aspects of the NUDGES framework that may also be important to the budget simulation choice architecture context, and we discuss the important theoretical distinction between our treatments and whether the treatment was a nudge or a budge (Oliver, 2015), which may be an important distinction for practicing choice architects. Beyond more than a wink and a nudge, we want practitioners to take behavioral analysis, nudges, and choice architecture seriously, and we encourage behavioral public policy researchers to explore possibilities for research in this area.

Choice architecture and budget simulations

The context of public budgeting is an important choice architecture environment that has stimulated foundational work in both behavioral theory and public

---

[2]'Unacceptably high' is the normative way of saying that these final budgetary surpluses would not be allowed by state law.

[3]This choice framing is important as government budgeters are encouraged to see themselves more as choice architects who gently nudge participants into fiscally responsible choices rather than just always tell them what can and cannot be done (Kavanagh and Argarunova, 2022). The research question is also embedded in a larger research question within the field of public budgeting: 'How can we use the information that is collected during budget engagement efforts to inform policy?' Without paying attention to important rules like balanced budget requirements, the information that is generated in these efforts may be interesting but also useless to the policy makers that have the responsibility to set the budget according to law. How to make people aware of regulatory requirements but also respect their autonomy to make decisions is overlooked in both research and practice.

policy practice. For example, Herbert Simon developed his foundational understanding of bounded rationality from a municipal budget setting process (Simon, 1992). Other research has looked at how budget presentation can influence perceptions and outcomes (McIver and Ostrom, 1976). Municipal budgeting now has tools for citizen outreach and engagement, and budget offices are excellent environments for research and choice architecture practice. Despite the wonderful potential of budgeting for studies of choice architecture, the choice architecture of public budgeting may remain underexplored because budgeting is an environment with laws that constrain choice. So, some nudges may lead people to make choices that conflict with what is allowed. We explore the possibility of nudging or budging people back into making choices that do comply with what is allowed. Since these are legal requirements, we use a stronger form of nudging, or budging (Oliver, 2013, 2015), to try to get people to comply with the budget laws. This also allows us to explore different causal mechanisms for why simulation participants expressed an interest in a budget outcome that would not be allowed by state law.[4]

The public sector, especially at the local level, has been increasing efforts to actively engage their residents. Citizen engagement is considered a best practice by groups such as the GFOA and the International City/County Management Association. There are many reasons for these citizen engagement efforts, including educating residents, increasing trust, and soliciting information on citizen preferences (Ebdon and Franklin, 2004; Afonso, 2021). While these efforts are being made throughout government, the budget presents a natural avenue for engagement since it lays out the scope of governmental activities, the costs and trade-offs of those choices, and represents a point where citizens may voice their preferences to influence outcomes and decisions.

Online budget simulations are increasingly popular virtual tools that allow participants, typically citizens of the jurisdiction, to take on the task of managing the government's budget. They are usually flexible instruments where public officials can define the spending categories and revenue options and set the starting levels. In some cases, the government may only allow participants to adjust expenditures or revenues, but many modern tools, such as Balancing Act, allow for dynamic environments where participants can (1) adjust both expenditures and revenues and (2) are asked to create a balanced budget where expenditures cannot exceed revenues. While online budget simulations originated at the national level, the majority of them are now at the local level.[5] In fact, over 100 municipalities in Canada have online budget simulations hosted by Citizen Budget and over 130 local governments in the United States have online budget simulations hosted by Balancing Act (Ethelo, 2019; Mohr and Afonso, 2024).[6] Among the reasons for implementing an online budget simulation is that it

---

[4]However, we acknowledge that our strong budge to get people to comply with the law may limit the theoretical findings. We discuss these limitations further in the 'Discussion' section.

[5]Where balance budgets are required.

[6]Not only has citizen engagement become a best practice, but online budget simulations have also been highlighted. The recipients of the Government Finance Officers Association's Award for Excellence in Public Engagement on the Budget in 2018, 2019, and 2020 were local governments with online budget simulations. Similarly, the 2020 recipient of the National Association of Counties Achievement Award in Financial

serves two of the primary purposes for citizen engagement. First, it is a tool for education because it allows participants to understand the scope and cost of government, and it shows them where their tax dollars are going. Second, it can also be a tool for information gathering. The government can select the categories that participants are able to modify, and the tool allows the government to collect and consider the responses.

While the trend toward greater use of online budget simulations is being led by budget practitioners, the academic community has begun to analyze how the choice architecture of the simulations impacts engagement and outcomes like budgetary preferences (Tuxhorn *et al.*, 2019, 2021, 2022; Mohr and Afonso, 2024). Tuxhorn *et al.* (2022) examined a federal budget using the Balancing Act budget simulation software using both a dynamic revenue and expenditure approach compared with the singular approach of stating revenue choices first and then expenditure choices. They found that the budgetary preferences in the dynamic simulation were significantly different than the preferences in the singular simulations. Likewise, Mohr and Afonso (2024) worked with practitioners to conduct a Balancing Act experiment (with dynamic feedback for all users) on a city's budget. They varied the starting conditions of budget balance, budget deficit, and budget surplus and used students as their sample. They found that changing the default starting value of the simulation influenced the budget choices and engagement with the simulation. Also, starting the budget simulation in a surplus as the default option led to a much higher end of simulation surplus or carryover, which would not be allowed by state law.[7]

The nudge framework in the choice architecture of budget simulations

Both academics and practitioners in the policy and management world have caught on to the importance of choice architecture as a research concept. According to Mannix and Dudley (2015, p. 711), the insights from choice architecture and nudge theory are '…useful in counseling people to make better decisions, including by designing government programs that provide information or present information in an accessible way'. The literature often recognizes the original heuristics and biases described in *Nudge*, in which Thaler and Sunstein (2009) made the case for choice architecture. For example, Dudley and Xie (2020) discuss the choice architecture in the case of regulation and look at the institutional factors that reform, mitigate, and aggravate four heuristics that they identify in the regulatory process: availability bias, confirmation bias, narrow framing, and overconfidence. Beyond heuristics and biases, Thaler and Sunstein (2009) define the six elements of choice architecture in terms of the acronym NUDGES (iNcentives, Understand mappings, Defaults, Give feedback, Expect error, Structure complex choices). These six elements go significantly beyond the heuristics and

---

Management went to Baltimore County, MD, for the strategic use of an online budget simulation to address a substantial deficit.

[7]While scholars may disagree about the wisdom of balanced budget requirements (Douglas and Raudla, 2024), it is a regulation that must be obeyed by the budget officers. Interestingly, municipal budget simulations do not generally mention these requirements, and this is all the more troubling, given that some nudges that intend to get people to reveal more budget preferences may lead to higher ending budget simulation surpluses (Mohr and Afonso, 2024).

biases approach and develop a comprehensive framework for understanding choice architecture.

The first element of the NUDGES framework is *incentives*, which includes both rational incentives and the psychological heuristics and biases that either hide or obscure rational incentives such as present bias and status quo bias. The manner in which choices are framed (Kahneman and Tversky, 1984) is likely to influence value trade-offs and how people respond.[8] In framing choices, *default choices* have been shown to be exceptionally important. Thaler and Sunstein note, 'people will take whatever option requires the least effort or the path of least resistance' (Thaler and Sunstein, 2009, p. 85).

The power of framing and defaults influences our first set of hypotheses. When starting conditions of the simulation in deficit or surplus, the outcomes will strongly be influenced by behavioral heuristics such as status quo bias (Samuelson and Zeckhauser, 1988), and defaults (Thaler and Sunstein, 2009). Since the simulation participant would not need to make changes in the simulation if they start in surplus due to status quo bias, behavioral theory would predict that starting in surplus would lead to a higher final balance. Likewise, this default option may also lead them to not make as many changes and the size of their changes will be smaller.

H1a: Final budget balance will be higher when starting from surplus.

H1b: The number of budget categories changed will be less when starting the simulation from a surplus.

H1c: The average size of the budget changes will be smaller when starting the simulation from a surplus.

Findings from the Mohr and Afonso study (2024) showed that many simulations users ended their simulation with large budget surpluses. Large ending budget surpluses in the budget simulations suggest overcorrection or possibly errors. In nudge theory, there is an interesting puzzle here. Are the budget respondents simply making a mistake or is there a behavioral reason for this overcorrection? For example, if they exhibit high ending surpluses when starting out with a budget in surplus, the default may have primed them to believe that surpluses may be good. However, the choice architecture framework and nudge theory would say that the budget simulation participants may have made an error or were using a simplifying strategy for a complex choice. Mohr and Afonso (2024) find evidence of a simplifying strategy of 'chunking' or making larger average changes when in the budget deficit condition. If these surpluses are not genuine policy preferences but simplifying strategies or chunking and

---

[8]Survey research has also come to similar conclusions. How we structure information influences choice outcomes (Achen, 1975; Zaller & Feldman, 1992). It is interesting that the relatively more complex simulation environment is likely to be an important evolution in the technology to elicit preferences. What is abundantly clear from this research is that the choices about how to present and structure information, or the choice architecture in behavioral policy terms, only becomes more important with this added complexity.

we were to warn the participants that their final budget simulation balance is too high, we would expect:

H2a: Simulation participants that are warned that their final budget surplus is too high will make additional changes to reduce their budget surplus.

H2b: Simulation participants that are warned that their final budget surplus is too high will increase the number of budget categories that are changed.

H2c: Simulation participants that are warned that their final budget surplus is too high will reduce the average size of their changes.

Finally, we believe there will be an interaction effect between the starting simulation condition and the ending surplus treatment. If a person starts in the surplus budget condition that is likely to lead to a higher ending budget simulation surplus, then the treatment to reduce the final simulation balance condition should interact with the starting condition to produce an even greater reduction in the budget surplus condition.

H3: Starting in surplus and getting an ending surplus treatment will interact to create an even larger reduction in the ending surplus.

These hypotheses point toward the importance of the choice architecture and the importance of nudges.[9] One could argue that the ending surplus treatment goes significantly beyond a nudge. While it preserves the essential autonomy of choice (Thaler and Sunstein, 2009) and is not a mandate, the treatment clearly relates the surplus to the important issue of legal compliance with budgetary laws. The distinction between different types of nudges has evolved since the original articulation of nudges by Thaler and Sunstein. For example, a nudge that is regulatory in nature but that is cognizant of behavioral economics approaches is considered a budge (Oliver, 2013, 2015). Typically, in online budget simulations the legal restrictions around ending surpluses would not be presented to participants. Therefore, while the ending surplus treatment does not ban large ending surpluses, it does go much further than what is observed in practice. Additionally, to be autonomy-preserving, we have allowed the participants to complete their budget simulations even if they do not get their final ending to perfectly balance or fall below the requirement set by state law.[10] This likely means that our treatment

---

[9]Thaler and Sunstein also suggest that good choice architecture also will understand people's welfare mappings, structure complex choices, and give feedback. While we have not tested these aspects of choice architecture, they can all be interestingly explored in the budget simulation environments. In the discussion, we discuss some ways that the budget simulations may allow researchers the ability to test welfare mappings, the optimal structure of complex choices, and options for feedback.

[10]Practitioners may want to be aware that a full budge that requires the simulation participant to get below the requirement may also be appropriate here, but they will need to weigh the tradeoff between a requirement and choice that will have implications for both the completion of the simulation and the revealed budgetary

is somewhere between a nudge and a budge, and so we simply call it the 'ending surplus treatment.' We discuss the distinction between nudges and budges further in the 'Discussion' section.

Research design

To address these hypotheses, we developed an experiment using a Balancing Act budget simulation and ran it in the fall of 2021. The simulation experiment was conducted by researchers at a large, urban research university in the southeastern United States using students as participants. This simulation was set up with the exact settings of the city's budget simulation. But the name of the municipality where the University is located was omitted at the request of city officials. Also, participants were randomly assigned to the different simulations. The students did the simulation as part of an online omnibus experiment[11] on their own internet-connected device, just as regular citizens of a community do.[12] Because of this setup, the experiment exhibits a high level of experimental and mundane realism (Iyengar, 2002) and the setup of the experiment would be nearly identical to the setup for the local government. However, the student sample is younger than typical citizen participants in the budget process (Maciag, 2014). A convenience sample (such as students) is not likely to have the same policy preferences as general citizen participants and is not representative of the broader populations' preferences. Thus, we are not suggesting that the substantive changes to the budget are representative of citizens' preferences, but the experimental treatments show the effect of different mechanisms for the testing and development of theory.[13] Therefore, we do not report on the policy preferences revealed when making changes to revenue or expenditure categories.

There are two benefits to local government in conducting these types of experiments with college students. First, we are able to assess the preferences of younger people that may not typically engage in these types of activities. Second, local governments are often risk averse and concerned about presenting imbalance to the general public, so the municipality upon whose budget this exercise is built was unwilling to present their budget to the public with either a surplus or a deficit. They did agree to allow us to present their budget to a student sample, to test whether starting participants in surplus or deficit and whether an ending surplus treatment before final submission influenced the hypothesized budget outcomes.

---

choices. This is further evidence of the craft and skill of the practicing choice architect in the budgetary environment.

[11]Students received extra credit in one of their courses for participating in the omnibus.

[12]We use *citizen* here to mean the members of a local community (Cooper and Gullick, 1984). We mean everyone that belongs to the community that can participate in the budget simulation.

[13]For example, Barnes *et al.* (2022) find that younger populations prefer lower taxes and spending than their older counterparts. There is also evidence that college age participants may have lower emotional response inhibition than older participants, making them more sensitive to behavioral nudges (Waring *et al.*, 2019) and older adults may have a processing bias towards positive information treatments (Reed *et al.*, 2014). Thus, scholars and practitioners must consider the possibility that the impact of the treatments may differ by age range; however, we do not believe that the treatments being tested here will differ meaningfully by age based on the existing literature.
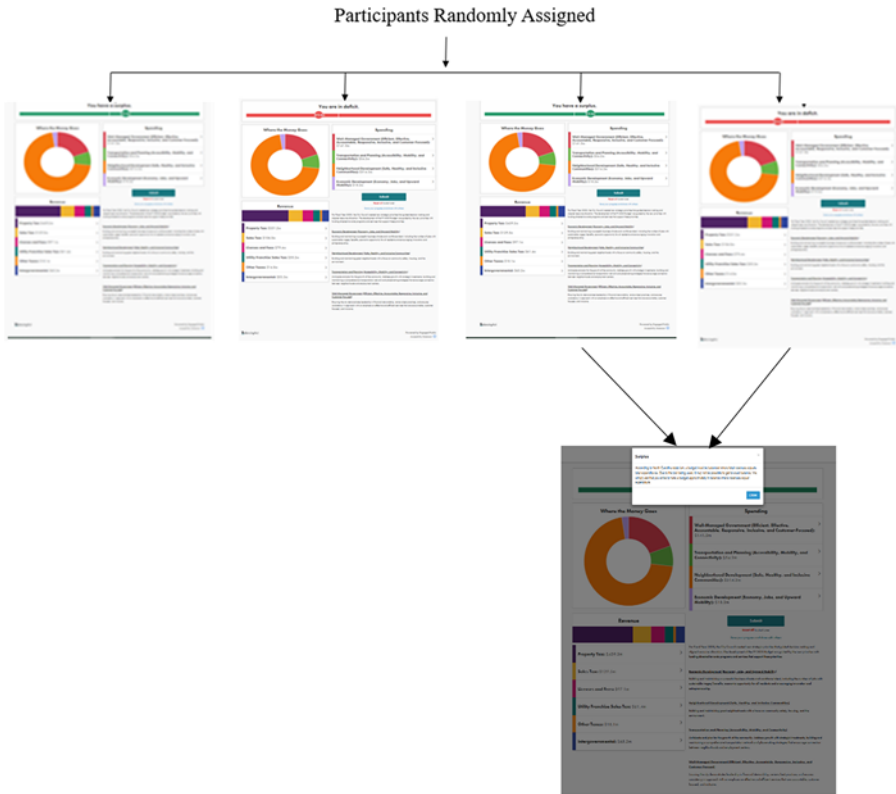
Figure 1. Participants randomly assigned to one of four conditions.

The participants were randomly assigned into one of four simulations:

1. Starting deficit with no ending surplus treatment,
2. Starting deficit with an ending surplus treatment,
3. Starting surplus with no ending surplus treatment, and
4. Starting surplus with an ending surplus treatment (Figure 1).

No matter the simulation assigned, all participants began with the same levels of spending, which is equal to the simulation used by the local government.[14] However, for participants beginning in deficit (surplus) the revenues were all decreased (increased) by 10 percent. This led to a budget deficit (surplus) of $72.8 million for the city and is graphically shown at the top by the red (green) 'balance bar' on the tool. When the budget is not in a state of balance or surplus, the bar is red; the participants can not complete the simulation until the revenues are either increased and/or

---

[14]The local government originally started its budget in balance, which was particularly uninformative as most people tended to default to the balanced budget, and the local government received very little information of any value.

Table 1. Sample descriptive statistics[a]

	Obs	Mean	Std. dev.	Min	Max
Age	357	20.294	3.658	16	46
Race					
White	358	0.587	0.493	0	1
African A.	358	0.151	0.358	0	1
Hispanic/Latino	358	0.095	0.294	0	1
Asian	358	0.067	0.250	0	1
NPI and other	358	0.039	0.194	0	1
Two or more	358	0.061	0.240	0	1
Gender					
Male	358	0.447	0.498	0	1
Female	358	0.545	0.499	0	1
Other	358	0.008	0.091	0	1

[a]Treatments not significantly different than the sample mean.

the spending decreased (see also Tuxhorn *et al.*, 2021). To adjust levels, the participants click on a category of revenue or expenditure and increase the up or down arrow until the balance bar is green and then they can click the 'Submit' button in the center. For the surplus condition where the revenues exceed the expenditures, the budget is technically balanced at the start of the simulation and no changes are required. However, the bar can become red, and the simulation cannot be completed if the participant were to make changes to the budget to take it out of balance.[15]

If a participant was in the ending surplus treatment group and submitted their final budget with the revenues exceeding the expenditures by more than half a percent,[16] the simulation participant was presented with an information box that read 'According to North Carolina state law, a budget must be balanced where total revenues equals total expenditures. Due to the tool being used, it may not be possible to get to exact balance. We simply ask that you strive to have a budget approximately in balance where revenues equal expenditure,' to nudge or budge them to further adjust the revenues and/or expenditures and reduce the ending surplus. It is important to note that the participants did not have to follow this prompt and could still submit their budget so long as revenues met or exceeded expenditures.

As can be seen in Table 1, the characteristics of the students that conducted our simulation experiment do not match the general population of the United States or the municipality. As would be expected, they are younger on average and there are more females than males. As is found in this southern city, the sample is quite diverse with 58.7% of the sample being white, 15.1% African American, 9.5% Hispanic,

---

[15]This setting is also the same as that which was being used by the local government when we conducted the study.

[16]Half a percent = 0.5%. Since the budget revenue was just over $800 million the exact value for showing the informational nudge message was a balance greater than $4,002,207.

Table 2. Final budget outcomes by ending surplus treatment and starting condition

	Final revenue		Final expenditure		Ending surplus	
	Coefficient	Standard error	Coefficient	Standard error	Coefficient	Standard error
Ending surplus treatment	−12.20	68.114**	15.10	8.798**	−27.20	7.102***
Surplus starting condition	115.00	6.899***	85.60	8.726***	29.70	7.043***
Constant	558.00	11.607***	559.00	14.711***	−0.86	11.900

Note: ***p < 0.01, **p < 0.05, *p < 0.1 for one-tailed tests. Values in millions of dollars.

6.7% Asian, 3.9% Native American and Other, and people of two or more races accounting for 6.1% of the sample. In the four conditions, balance tests indicate that there are not significant differences between the four groups. The results are modeled with ordinary least squares (OLS) regression.

Analysis

We begin the analysis with the results of varying the starting condition and the ending surplus treatment on simulation participant behavior and budgetary choices and the impact of the choice architecture on whether either of these interventions actually led to decreases in ending surpluses (Table 2).[17] In previous work, there is evidence that starting condition influences ending surplus (Mohr and Afonso, 2024). Here, we seek to measure the impact of the ending surplus treatment while controlling for the two different starting conditions on behavioral responses from participants. To do this, we run OLS regressions modeling whether participants were in the ending surplus treatment group and their starting condition, the binary treatment variables, on our outcome of interest, ending surplus.[18]

The results (Table 2) suggest that both the ending surplus treatment and starting condition have a statistically significant (p < 0.01) and practically significant impact on the ending surplus. Similar to the Mohr and Afonso analysis (2024), we find that a starting condition of budgetary surplus increases the ending surplus by over $29 million, supporting *H1a*. We also find evidence that the ending surplus treatment reduces the ending surplus by a similar magnitude, a reduction of over $27 million, supporting *H2a*.

Table 2 also presents the impact of starting condition and the ending surplus treatment on final revenues and expenditures. While we do not hypothesize on how participants will adjust their budgets in the simulation regarding revenues and expenditures, we do find interesting results. Participants beginning the simulation in surplus

---

[17]In the appendix, we also present these results where we do not control for both treatments and only model the impact of starting condition and the ending surplus treatment.

[18]OLS regression is appropriate for a continuous outcome value like the budgetary outcomes we are modeling here. We use one-tailed tests because we have directional hypotheses and one-tailed tests provide greater power to detect an effect in the direction of our hypotheses.

Table 3. Number of changes by ending surplus treatment and starting condition

	Expenditures		Revenues		Total number of changes	
	Coefficient	Standard error	Coefficient	Standard error	Coefficient	Standard error
Ending surplus treatment	2.045	1.069**	0.602	0.181***	2.651	1.113***
Surplus starting condition	−1.523	1.059	−0.815	0.180***	−2.339	1.103***
Constant	11.824	1.786***	2.083	0.303***	13.906	1.859***

Note: ***p < 0.01, **p < 0.05, *p < 0.1 for one-tailed tests.

have, on average, \$115 million more in final revenues than those who begin in deficit ($p < 0.01$) and \$85 million more in expenditures ($p < 0.01$). Since both starting conditions begin with an equal level of expenditures and a difference in revenues of close to \$146 million, this suggests participants presented a budget where expenditures do not equal revenues tend to use basic budget balancing tactics of both reducing expenditures and increasing revenues. These results shed light on the differences in outcomes, and we can observe that the total size of the budget remains larger for those who begin the simulation in surplus since total expenditures and ending surplus are significantly higher.

Table 3 presents the number of expenditure categories, revenue categories, and total categories changed in the simulation, and offers support for *H1b* and *H2b*, meaning that both starting condition and ending surplus treatment impact their engagement with the tool. We find, in keeping with earlier analysis (Mohr and Afonso, 2024), that participants who begin the simulation in surplus are more likely to make fewer changes to the tool. In total, they are more likely to make 2.3 fewer changes than those who begin in deficit. They are also less likely to make adjustments to revenues, but there is no statistically significant impact on the number of expenditure categories changed. Similarly, we see that the ending surplus treatment positively impacts engagement with the tool as measured by the number of categories adjusted. We see this impact especially on revenues. Participants are likely to make 0.6 more changes, out of a possible 3, to revenues when treated with the ending surplus treatment ($p < 0.01$) and just over 2 additional changes to expenditures ($p < 0.05$). Overall, they are more likely to make 2.6 more changes to expenditure and revenue categories when presented with the ending surplus treatment ($p < 0.01$).

Another way of considering engagement with the tool is to understand what the absolute value of the changes to the categories are. This is an important consideration because a primary goal of a tool such as online budget simulations is for participants to reveal their preferences around local policies. Thus, a \$2 million decrease to one area of expenditure coupled with a \$2 million increase to another will not be revealed in aggregate changes but does show engagement by the participant and that the host governments are receiving feedback on participant policy preferences. Given our student sample, we do not believe that the policy preferences are likely representative, but the way the treatments impact engagement are representative. We find evidence for *H1c* and *H2c*, reported in Table 4, that both the starting condition and the ending

Table 4. Changes in absolute value by ending surplus treatment and starting condition

	Expenditures		Revenues		Total change	
	Coefficient	Standard error	Coefficient	Standard error	Coefficient	Standard error
Ending surplus treatment	−2.71	12.70	11.90	7.146**	−6.72	3.744**
Surplus starting condition	−34.80	12.632***	−8.64	7.09	−8.77	3.713***
Constant	138.00	21.221***	31.40	11.944***	26.30	6.257***

Note: ***p < 0.01, **p < 0.05, *p < 0.1 for one-tailed tests. Values in millions of dollars.

surplus treatment impact engagement with the tool as measured by the absolute value of changes. As expected, participants beginning in surplus make smaller changes in the absolute value of the expenditures by almost \$35 million (p < 0.01) and the total changes by \$8.8 million (p < 0.01). We find no statistically significant impact on the absolute values of the changes to revenues. The ending surplus treatment increases the absolute value of changes to revenues by almost \$12 million (p < 0.05) and the absolute value of the total changes by over \$6.5 million (p < 0.05). Unlike for starting condition, we find no evidence that the ending surplus treatment impacts the absolute value change on expenditures. This may be because participants reduced the magnitude of their initial changes, thus negating the end result. Ultimately, we find support for our hypotheses and evidence on how the choice architecture impacts engagement with online budget simulations as measured by adherence with law and practice (reducing ending surplus), the number of changes made to the budget, and the absolute value of those changes.

Table 5 presents the results of a regression that controls for both whether the participant began the simulation in a surplus rather than a deficit, as well as whether they received the ending surplus treatment, and the interaction between the two treatments. The results suggest that participants beginning the exercise in surplus will have a larger ending surplus, by \$38 million, than those who begin with a deficit (p < 0.01). This is in keeping with the findings presented in Table 2. We continue to find that the ending surplus treatment reduces ending surplus by approximately \$16.4 million (p < 0.1). Similarly, the interaction term between beginning in surplus and ending surplus treatment is also negative, and participants that receive the ending surplus treatment and

Table 5. Final budget outcome by ending surplus treatment and starting condition with an interaction

	Ending surplus	
	Coefficient	Standard error
Ending surplus treatment	−16.40	10.200*
Surplus starting condition	38.80	9.329***
Interaction	−21.10	14.211*
Constant	23.80	6.919***

Note: ***p < 0.01, **p < 0.05, *p < 0.1 for one-tailed tests. Values in millions of dollars.

begin in surplus reduce their ending surplus by an additional $21 million, on average, (p < 0.1). Therefore, we find evidence for *H3*, though it does not meet the standard criteria for statistical significance. The interaction terms are not statistically significant for the models where final revenues and expenditures are the outcomes of interest, suggesting that the combination of choice of starting condition and ending surplus treatment do not interact.

Discussion

When looking at overall engagement with the tool, our results partially replicate earlier findings and show that the final ending surplus treatment impacts engagement with the simulation in meaningful ways. We find that the final ending surplus treatment impacts the behavior of reducing the ending surplus. Further analysis suggests that the most common adjustments made to reduce the ending budgetary surplus are reductions to revenue. This is an interesting finding that is worth exploring in future research. It may suggest that participants have a stronger preferred level of expenditures relative to revenues consistent with Tuxhorn *et al.* (2022), and it may be a revealed preference unique to this population because the students may be less sensitive to levels of taxation than the general population. However, once we control for the starting condition, we find that the impact of the ending surplus treatment is to reduce the ending surplus, as intended, and that it does not dramatically change other forms of engagement. This suggests that if practitioners are interested in online budget simulations and want to reduce the noise of large ending surpluses, they can implement an ending surplus treatment to participants and still receive high levels of engagement. In the case of *H3* (the interaction hypothesis), theory suggests that the explanation for the large ending surpluses is likely attributable to either loss aversion which would be intentional or because of errors likely caused by simplifying strategies. The ending surplus treatment may impact the ending surplus levels, but we would expect the ending surplus treatment to have a larger impact when the ending surplus was the result of errors than whether it was an intentional choice caused by loss aversion. The results presented in Table 5 suggest that high ending surpluses are the result of simplifying strategies and possibly error.

This research makes contributions to nudge (and budge) theory and choice architecture by extending the theory into the complex decision environment of public budgeting. Using dynamic online budget simulations, we were able to test key propositions of nudge theory and examine two choices that have to be made in structuring the online budget simulation. We add to the literature that has shown that giving feedback changes outcomes (Tuxhorn *et al.*, 2022) and that defaults and behavioral incentives also change budgetary outcomes (Mohr and Afonso, 2024). In this paper, we examined the theoretical dimensions of expecting errors through the impact of an informational intervention, which led to an outcome that is more realistic relative to what would be allowed by state law and practice.

This project links nudge theory with behavioral budget and finance to examine dynamic online budget simulations. While the literature has now begun to examine defaults, giving feedback, incentives, expecting errors, and structuring complex choices, much more work could be done to understand these issues. First, this research

points to the need to understand further how to structure complex choices. We used the original categories presented in a municipal budget simulation; but perhaps simplifying these categories more broadly into seven or eight larger categories could help people structure their own thinking about preferred trade-offs. Second, the default levels were based off Mohr and Afonso (2024) and these levels may be different in real life budget scenarios where starting conditions may be less (or more extreme).

More could also be done with the impact of feedback and understanding welfare mappings on simulation outcomes. For example, what does it mean if we cut the police budget by 10 percent? What will that do to service levels? In the abstract, making a cut like that may seem attractive because the police budget category is large, which can help the person achieve balance quickly. However, if the person were presented performance information about how increases or decreases are likely to affect the service and affect welfare, then we suspect that this may strongly influence the budget choices. Understanding these complex choice environments further can help develop nudge theory further in this area and improve budgetary practice.

The practical implications of this work are numerous. First, the choice architecture and how the simulation is structured impacts the budget outcomes revealed by participants and the amount of engagement. The budget prompt nudging participants to minimize ending surplus led to greater engagement, but it did not impact the choice of budget balancing strategies. Second, the choice architecture also impacted budget outcomes in the simulation. Ending the simulation with a prompt to encourage participants to not carry a large ending surplus accomplishes this goal. In the experiment, it cut the expected ending surplus by two thirds.

These modifications to the choice architecture are meaningful for practitioners. The goals of citizen engagement efforts, such as online budget simulations, typically revolve around (1) education or information sharing with citizens, (2) consulting with citizens on specific issues, or (3) actively partnering with citizens to shape policies (Afonso, 2021). Online budget simulations can be powerful tools for both education and information sharing and consultation with citizens. Online budget simulations present budget information in a digestible format where participants receive a high-level snapshot on the scope, cost, and areas of expenditure of their government, as well as how it is financed. The tool can also act as consultation, where participants provide feedback on preferences for specific policy options or the entire budget. Online budget simulations let governments learn where residents would like to spend more money, where they will tolerate reductions, and their preferences for revenue policies within a framework where participants have to balance the budget while not just signaling that they prefer increased expenditures and decreased revenues without acknowledging the balance. Given that governments want to maximize education and consultation goals, increasing engagement with the tool and better capturing accurate preferences by reducing errors is important. As with all citizen engagement efforts, it is critical for governments to carefully consider their goals for the engagement when choosing and structuring tools. This research helps inform their strategies once their goals are set.

However, this is not to suggest that there will not be barriers to implementing these changes. For example, it is possible that the implementation of an ending surplus treatment may encourage participants, who have revealed their true preference

for an ending surplus, to modify their response. In the case of a large ending surplus, practitioners may want to consider allowing participants to increase fund balance or savings. This would be in keeping with the law, but also allow for the possibility of loss aversion. More broadly, these budges that influence what may be errors or revealed preferences may be appropriate, e.g. if participants' preferences are illegal or infeasible, but in other cases it may not be aligned with the goals of the community. Therefore, we encourage practitioners to carefully consider any intervention and its wording/execution and how it may impact responses in both desired and undesired ways. A broader concern may be that public officials are reluctant to implement starting conditions that are not in balance. There may be concerns over appearing to have overtaxed residents when starting in surplus or mismanaging funds when starting in deficit. While these are genuine concerns, the results of an earlier experiment suggest that trust in government was not impacted by starting condition (Afonso *et al.*, 2023). Furthermore, practitioners can look to cities like Charlotte, NC. and Lawrence, KS. who as of this writing begin their online budget simulations in deficit based on this stream of research.

This research shows that governments can nudge or budge participants to make their online budget simulations more relevant and meaningful. Practitioners may want to further explore work on different classifications of nudging and what these might look like in practice (Oliver, 2013, 2015). Oliver (2015) has developed a useful taxonomy of nudges, budges, and shoves. He maps rational-behavioral responses, responses with more internalities or externalities, and nudges that are either more concerned with regulation or liberty. Others have promoted the power of boosts (van Roekel *et al.*, 2022) and deliberation (Banerjee and John, 2024). While we have been mostly concerned with the difference of regulatory-preserving vs liberty-preserving nudges, it may be especially useful to design public budgeting experiments that trigger more rational responses vs behavioral ones to further understand welfare mapping of internalities and externalities.

The research also has notable limitations. Foremost, the respondents are students, with responses unlikely to be representative of the broader population, even though broader populations (as theory suggests) would have answered in much the same way as the students. A second important consideration is that the wording of our ending surplus treatment may have been interpreted less as a nudge and more as an instruction. Particularly, our finding that the high ending surplus is being driven more by respondent error in the simulation rather than loss aversion may be driven by this treatment if it is too overpowering and respondents *feel* like they have no choice. Further research on simulation design, similar to survey design (i.e. Achen, 1975; Zaller and Feldman, 1992), is likely to be an emerging area of scholarship and practice that we want to encourage.

Conclusion

Choice architecture is more than just incentives and simplifying heuristics that get used by boundedly rational people. In thinking about choice architecture, scholars and practitioners can use the NUDGES framework to think about how to structure complex choices, understand people's mappings and explain the choice to get them to

consider how the choice may influence their welfare, understand that people are going to make errors, give feedback, and be mindful of the default and easy choices. By more thoroughly integrating the theory of choice architecture into public management and public policy studies, we gain more theoretical leverage over the ways that people reveal preferences and make policy choices.

Beyond nudges, practicing choice architects may also need to think about regulatory requirements that may require more of a budging approach. While the approach that we have developed here may not be a perfect nudge or budge, there is value in showing practitioners this development. As they try to build more realistic budget simulations, they will be confronted with the challenges of both revealing the public's preferences but also doing it in a way that is consistent with the legal and practical environment that is also an important part of their job.

References

Achen, C. H. (1975), 'Mass political attitudes and the survey response', *American Political Science Review*, **69**(4): 1218–1231.

Afonso, W. (2021), 'Citizen engagement through the budget process', in W. Afonso (ed), *Budgeting in North Carolina Local Governments*, 2nd edn, Chapel Hill: University of North Carolina Press, 303–322.

Afonso, W., Z. Mohr and S. Powell (2023), 'Citizen Engagement Tools: online Budget Simulations', *Death & Taxes*, University of North Carolina. https://deathandtaxes.sog.unc.edu/citizen-engagement-tools-online-budget-simulations/ [11 January, 2024].

Banerjee, S. and P. John (2024), 'Nudge plus: incorporating reflection into behavioral public policy', *Behavioural Public Policy*, **8**(1): 69–84.

Barnes, L., J. Blumenau and B. E. Lauderdale (2022), 'Measuring attitudes toward public spending using a multivariate tax summary experiment', *American Journal of Political Science*, **66**(1): 205–221.

Castleman, B. L., F. X. Murphy, R. W. Patterson and W. L. Skimmyhorn (2021), 'Nudges don't work when the benefits are ambiguous: evidence from a high-stakes education program', *Journal of Policy Analysis and Management*, **40**(4): 1230–1248.

Cooper, T. L. and L. Gulick (1984), 'Citizenship and Professionalism in Public Administration', *Public Administration Review* 44: 143–151.

Douglas, J. W. and R. Raudla (2024), 'Do balanced budget practices of US states make sense? Alternatives from the Eurozone', *Journal of Public Budgeting, Accounting & Financial Management*.

Dudley, S. E. and Z. Xie (2020), 'Designing a choice architecture for regulators', *Public Administration Review*, **80**(1): 151–156.

Ebdon, C. and A. Franklin (2004), 'Searching for a role for citizens in the budget process', *Public Budgeting and Finance*, **24**(1): 32–49.

Ethelo. (2019), 'Ethelo Acquires Citizen Budget', https://blog.ethelo.org/press-release-ethelo-acquires-citizen-budget.

Iyengar, S. (2002), *Experimental designs for political communication research: from shopping malls to the internet*. Paper presented at the Workshop in Mass Media Economics, Department of Political Science, London School of Economics.

Kahneman, D. and A. Tversky (1984), 'Choices, values, and frames', *American Psychologist*, **39**(4): 341.

Kavanagh, S. and S. Agarunova (2022), Rethinking Budgeting: Are We Ready for a New Approach? PM Magazine, International City/County Management Association.

Linos, E., L. T. Quan and E. Kirkman (2020), 'Nudging early reduces administrative burden: three field experiments to improve code enforcement', *Journal of Policy Analysis and Management*, **39**(1): 243–265.

Maciag, M. (2014), 'The Citizens Most Vocal in Local Government', *Governing*. https://www.governing.com/archive/gov-national-survey-shows-citizens-most-vocal-active-in-local-government.html [11 January, 2024].

Mannix, B. F. and S. E. Dudley (2015), 'The limits of irrationality as a rationale for regulation', *Journal of Policy Analysis and Management*, **34**(3): 705–712.

McIver, J. P. and E. Ostrom (1976), 'Using budget pies to reveal preferences: validation of a survey instrument', *Policy & Politics*, **4**(4): 87–110.

Mohr, Z. and W. Afonso (2024), 'Budget starting position matters: a "field-in-lab" experiment testing simulation engagement and budgetary preferences', *Public Budgeting & Finance*, **44**(1): 38–59.

Mohr, Z. and L. Kearney (2021), 'Behavioral-experimental public budgeting and financial management: a review of experimental studies in the field', *Public Finance and Management*, **20**(1): 11–44.

Oliver, A. (2013), 'From nudging to budging: using behavioural economics to inform public sector polic', *Journal of social policy*, **42**(4): 685–700.

Oliver, A. (2015), 'Nudging, shoving, and budging: behavioural economic-informed policy', *Public Administration*, **93**(3): 700–714.

Reed, A. E., L. Chan and J. A. Mikels (2014), 'Meta-analysis of the age-related positivity effect: age differences in preferences for positive over negative information', *Psychology and Aging*, **29**(1): 1.

Riis, J. and J. Peterson (2022), *Budget Officer as Decision Architect*. Rethinking Budgeting, Government Finance Officers Association.

Samuelson, W. and R. Zeckhauser (1988), 'Status quo bias in decision making', *Journal of Risk and Uncertainty*, **1**(1): 7–59.

Simon, H. A. (1992), 'Rational decision-making in business organizations', *Economic Sciences (1968-1980). The Sveriges Riksbank (Bank of Sweden) Prize in Economic Sciences in Memory of Alfred Nobel*, **1**: 343–371.

Thaler, R. H. and C. R. Sunstein (2009), *Nudge: Improving Decisions about Health, Wealth, and Happiness*, New York: Penguin.

Thaler, R. H. and C. R. Sunstein (2021), *Nudge: The final edition*, Yale University Press.

Tuxhorn, K. L., J. D'Attoma and S. Steinmo (2019), 'Trust in government: narrowing the ideological gap over the federal budget', *Journal of Behavioral Public Administration*, **2**(1): 1–13.

Tuxhorn, K. L., J. D'Attoma and S. Steinmo (2021), 'Do citizens want something for nothing? Mass attitudes and the federal budget', *Politics & Policy*, **49**(3): 566–593.

Tuxhorn, K. L., J. D'Attoma and S. Steinmo (2022), 'Assessing the stability of fiscal attitudes: evidence from a survey experiment', *Public Administration*, **100**(3): 633–652.

van Roekel, H., J. Reinhard and S. Grimmelikhuijsen (2022), 'Improving hand hygiene in hospitals: comparing the effect of a nudge and a boost on protocol compliance', *Behavioural Public Policy*, **6**(1): 52–74.

Waring, J. D., T. R. Greif and E. J. Lenze (2019), 'Emotional response inhibition is greater in older than younger adults', *Frontiers in Psychology*, **10**: 961.

Zaller, J. and S. Feldman (1992), 'A simple theory of the survey response: Answering questions versus revealing preferences', *American Journal of Political Science*: 579–616.