BEHAVIORAL-EXPERIMENTAL PUBLIC BUDGETING AND FINANCIAL MANAGEMENT: A REVIEW OF EXPERIMENTAL STUDIES IN THE FIELD

Zachary Mohr, PhD CGFM
Department of Political Science and Public Administration
The University of North Carolina at Charlotte

Linwood Kearney, PhD CPA
Department of Accounting
The University of North Carolina at Charlotte

ABSTRACT

Behavioral research in public budgeting and financial management (PBFM) has a venerable lineage that extends all the way back to Herbert Simon's early analysis of budgetary processes. In spite of this early connection to public budgeting, behavioral research is now more commonly associated with behavioral economics, psychology and other disciplines. Importantly, the term behavioral is now most often associated with the use of experimental research designs. We argue that the common perception in PBFM is that there is not a significant amount of behavioral-experimental (B-E) research currently being conducted in PBFM. We find that contrary to this expectation that there are some important areas where PBFM researchers have used experiments extensively including evaluating budgetary tradeoffs, the impact of performance information on budgeting outcomes, contracting and purchasing, and accountability. We then review related literatures for studies in other fields that are related to areas of PBFM research. From this review of the literature on experiments in PBFM and related fields, we see significant differences in both theoretical development and experimental designs. In conclusion, we suggest that for Behavioral-Experimental PBFM (B-E PBFM) to flourish that the field should be open to new theories and experimental methods.

Keywords: experiment, quasi-experiment, budgeting, financial management, accounting

1. INTRODUCTION

Behavioralism, or the study of actual human behavior as opposed to behavior theorized by neoclassical economics, began with studies of public budgets (Ridley and Simon, 1937; Simon, 1947).[1] Since that time, however, behavioral research,

1. See McCaffery and Slemrod (2006, p.3) for a similar definition for behavioral public finance. To achieve more practical and real-world relevance, they say that behavioral public finance needs to incorporate behavioral insights from psychology and behavioral economics. Thaler (2015) acknowledges the work of Simon as the start of behavioralism. However, he notes that the field of behavioral-economics did not begin in earnest

and especially behavioral experiments, have been dominated by other disciplines like psychology, sociology, management, and more recently by economics. Increasingly, behavioral studies have come to mean the use of experiments or quasi-experiments to study the behavior of people and groups of people (James, Jilke, and Van Ryzin, 2017).[2] While it is commonly assumed that public administration in general and public budgeting and financial management (PBFM) in particular has not been extensively doing experimental behavioral research for very long, there are areas of public budgeting that have been doing experiments for a fairly long time (i.e. Arrington and Jordan 1982). These studies are a strong base upon which behavioral experiments in public budgeting and financial management can be developed. Additionally, the literature outside of the field of PBFM may also be built upon to develop the theory and methods for experimental PBFM research.

In this paper, our research questions are 1.) What areas of PBFM research have used experimental research designs extensively? 2.) What literature can we find in other fields that use experiments that relates to PBFM research? 3.) What can we learn from the development of the use of experiments in other fields that can guide the still nascent field of behavioral-experimental public budgeting and financial management? To address these questions, we first define behavioral-experimental public budgeting and financial management (B-E PBFM) research and then review the literature in PBFM and related areas. From these studies we see that experiments may be particularly strong for making causal inference and helping to develop testable theories for PBFM. However, the public setting presents additional challenges that are not often experienced in some fields like economics and psychology, which may allow them to limit their analysis exclusively to the decision-making process of individuals. We discuss some of the challenges that we see from our review of the literature, and we encourage researchers to pursue a variety of experimental methodologies to develop new and innovative B-E PBFM theories and methods.

---

until the foundational experiments of Kahneman and Tversky (1979). While we tend to view the field of behavioral budget and finance as older, and broader, than experimental budget and finance, it is important to note that the older behavioralism informs the newer behavioral-experimental research.

2. Experiments are defined here as research that involves a treatment that is randomly assigned. The dependent variable is then observed and evaluated for both the treatment and the control group. When the treatment is not assigned by true, random assignment, but may be determined to be nearly random assignment, then we refer to these types of experiments as quasi-experiments. A good example of a quasi-experiment is the use of election monitors who are trained to move between polling places as-if the polling locations are chosen at random and then evaluating the evidence of fraudulent voting in the sites that the election monitors visit and the sites that do not have election monitors (Hyde, 2007). In this article, both experiments and quasi-experiments are discussed. Studies that try to mirror the logic of experiments like difference in difference and regression discontinuity for purposes of causal inference, but do not have random assignment or nearly random assignment into a treatment are generally outside the scope of this paper.

2. GROUNDING BEHAVIORAL-EXPERIMENTAL PUBLIC BUDGETING AND FINANCIAL MANAGEMENT

Behavioralism, or the study of individual and group decision making, is often attributed to Herbert Simon (Grimmelikhuijsen, Jilke, Olsen, and Tummers, 2017; Simon, 1947, p.352). In the 1950s and 1960s, scholars such as Herbert Simon (1955), Dwight Waldo (1948), Frederick Mosher (1956), and Richard Cyert and James March (1963) conducted behavioral research. However, while they all indicated that decision making was central to administration and a need for greater integration of psychology into administrative decision making, they tended to use primarily formal theory and inductive processes to generate their insights. While behavioral research progressed to more formal qualitative and quantitative models in public administration,[3] the use of experiments in behavioral public administration has not been a significant research strategy until recently (Bouwman and Grimmelikhuijsen, 2016). We first trace the rise of behavioral studies in PBFM that use traditional, observational methods (studies using qualitative, quantitative, and mixed method studies). Then we argue that behavioral, as it is more recently used in behavioral economics and behavioral public administration, is primarily the use of experiments to test fundamental theories of decision making both at an individual and group level.

Much of the foundational budgeting research can be considered behavioral in the sense that bounded rationality and satisficing often lead to the budgetary decision processes we observe. While the writings of Aaron Wildavsky (1964) were not explicitly behavioral, they fit comfortably within the framework espoused by Simon. In short, Simon argued that, due to bounded rationality, people tend to satisfice in their decision making process and limit the choice set to ease decision making (Simon, 1947, 1955). Incremental budgeting can be seen as the choice between a marginal increase or decrease in the budget for services and very few large budgetary changes (Flink, 2018; Jordan, 2003). This is very different from the perfect rationality of public finance that Simon discussed in his Nobel lecture where the budget is determined based upon the marginal utility of individual items in the choice set (Simon, 1992). In his early work, he found a puzzling phenomenon that he describes as follows:

3. According to Frederickson and Smith (2003, p.180), studies of bounded rationality "…ordinarily use qualitative methods; case studies based on observations, interviews and surveys are a staple. Cases also sometime use quantitative data (Brehm, Gates, and Gomez, 1998). Synthesis combined with modeling, using that word in the sociological sense, are common (Lipsky, 1980; Yanow, 1996)…." In experimental terminology, these types of studies are often referred to simply as observational studies to distinguish them from experimental and quasi-experimental designs that have a treatment and control that are randomly assigned to the research subjects.

> "Although the heads of the two agencies appeared to agree as to the objectives of the recreation program, and did not appear to be competing for empire, yet there was continual disagreement and tension between them with respect to the allocation of funds between physical maintenance, on the one hand, and play supervision on the other. Why did they not, as my economics books suggested, simply balance off the marginal return of the one activity against that of the other? Further exploration made it apparent that they didn't equate expenditures at the margin because, intellectually, they couldn't." (Simon, 1992, p. 352).

From this puzzle, Simon developed the ideas of satisficing and bounded rationality, which may drive the incremental budgetary processes so often observed. Additionally, related puzzles and ideas are found throughout the literature of PBFM. Analysts of budgeting note the importance of individuals and individual decision making, such as Irene Rubin's thick description of how individual politics can intersect with policy and process (Rubin, 2008); or Kurt Thurmaier and Katherine Willoughby's policy and politics frames that budgeters have to consider (Thurmaier and Willoughby, 2014). The alternatives in the decision-making framework usually gets reduced to a manageable set in the decision-making process. Budget research also examines how specific tools can be used to increase budgetary rationality for individuals or departments (Ho, 2011), how process influences budget decisions (Rubin, 2008), and even how decreases in a budget influence people differently than increases (Behn, 1985; Levine, 1978). In short, a great deal of observational budgetary research is behavioral (perhaps even most), because it looks at the decisions and decision-making processes of individuals or small groups of individuals.

In financial management, there are also good examples of behavioral work. Miller (1991) discusses how the activities of government finance officers are often designed to reduce risk and uncertainty. Other researchers discuss behavioral implications in other public financial management areas like debt management (Justice and Miller, 2011; Miller, 1993). One of the key insights to this research is that rationality cannot be perfect and so processes become extremely important to overcome individual limitations and comport with democratic norms (McCue, 2000). In short, much of the analysis in public budgeting and a significant amount in financial management is behavioral in the sense of looking at decision making processes, primarily with the use of observational methods.

More recently, there has been a growth in the use of behavioral public administration to mean the use of experiments in public administration and public policy (Bouwman and Grimmelikhuijsen, 2016; Grimmelikhuijsen et al., 2017). Research

using experiments has grown exponentially in the last decade in public administration (Bouwman and Grimmelikhuijsen, 2016) in large part due to leading journals publishing more research, the emergence of new specialized journals like the *Journal of Behavioral Public Administration*, and because of many journals doing special symposia and special issues such as this one.

This experimental research has no doubt been encouraged by the field of behavioral economics having several Nobel laureates added to its ranks in recent years whose work is directly applicable to B-E PBFM. In 2002 Daniel Kahneman, a psychologist, was awarded the Nobel Prize in Economics[4] for his work on decision making heuristics and prospect theory (Kahneman, 2003; Kahneman and Tversky, 1979). In 2017, Richard Thaler won it for his work on non-rational decision making and its application to economics and public policy (Thaler and Sunstein, 2009). This last year, the work of Michael Kremer, Abhijit Banerjee, Esther Duflo on the use of experiments in development economics was also awarded the prize (Banerjee, Deaton, and Duflo, 2004; Kremer, 2003). This research is notable for its use of experiments for testing the limits of foundational theories and has led to new insights for both economics and applied public policy (John, 2017, 2018). It is notable that there have even been applied experimental research teams, such as the Behavioral Insights Team in Great Britain, the Social and Behavioral Sciences Team in the Obama administration, and now the Office of Evaluation Sciences in the General Service Administration in the U.S federal government.

The key distinction that we make between the traditional behavioral research and the newer research is that the newer behavioral-experimental (B-E) research primarily uses experiments to make stronger claims for causal inference. As Richard Thaler (2015) notes, this research starts after Kahneman and Tversky (1979; Kahneman 2003) and uses experiments almost exclusively. While panel and time series data may also allow causal inference, it is often difficult to collect individual level or small group level data over enough time periods to make causal inference. Therefore, B-E research uses experiments to make causal inference to individuals and small groups.

4. Technically, the Nobel Prize in Economics is not one of the original Nobel Prizes established by Alfred Nobel, but is Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel. However, as is commonly accepted practice we simply refer to it as the Nobel Prize in Economics. Vernon Smith was also awarded the Nobel prize in 2002 for his experimental work, but since his research has had primary application to micro-economic theory and not broader policy, his work is not considered extensively here. Amos Tversky died prior to prospect theory being recognized for the Nobel prize in economics. His work on prospect theory and perceptions of risk are foundational in behavioral economics and directly applicable to B-E PBFM.

While there are no explicit statements on the use of experiments in the field of PBFM, the perception of the use of experiments in PBFM research is probably fairly similar to the field of public administration that did not have many published experimental studies until recently (Li and Van Ryzin, 2017). While there were clearly studies being done by behavioral economists in public finance (McCaffery and Slemrod, 2006), the budgeting experiments done by public budgeting and financial management scholars are not nearly as well known. Therefore, we believe that the common *perception* is that there are not many published experiments in budgeting and financial management.

3. PUBLIC BUDGET AND FINANCIAL MANAGEMENT EXPERIMENTS

While the number of experiments in PBFM research is not often recognized, we have found through our research that there are many experiments in the PBFM literature (Table 1). First both chronologically and in terms of its development, research into budgetary tradeoffs has long used experiments to test its hypotheses. Research in public sector contracting, performance measurement, and accountability also have several published articles that use experimental research designs. These experiments are a good base upon which to build B-E PBFM and it suggests the current level of development of experimental research in PBFM.

Table 1. Experiments Discussed in PBFM Literature

Author (Date)	Paper Title	Exp. Type	Experiment/ Quasi-Experiment	Sample/Unit of Analysis
Arrington, T. S., & Jordan, D. D. (1982)	Willingness to pay per capita costs as a measure of support for urban services.	Survey	Single factor experiment - control no tax cost and treatment is average tax cost	150 random Charlotte NC voters
Baekgaard, M. (2015)	Performance Information and Citizen Service Attitudes: Do Cost Information and Service Use Affect the Relationship?	Survey	2(Performance information) x 2 (cost information)	1866 Danish respondents in internet panel
Baekgaard, M., Serritzlew, S., & Blom-Hansen, J. (2016)	Causes of Fiscal Illusion: Lack of Information or Lack of Attention?	Survey	2 Experiments - vary cost framing and level of cost in each	Exp 1 - 1141 Danish respondents; Exp. 2 1839 Danish respondents to internet panel

Borry, E. L., et al. (2018)	Formalization and consistency heighten organizational rule following: Experimental and survey evidence.	Survey	2 experiments: #1 varies rule formalization, #2. 2(rule formalization) x 2(rule consistency)	Exp 1 -74 MPA students; Exp 2 - 150 public works managers
Brunner, E. J., Robbins, M. D., & Simonsen, B. (2018)	Information, Tax Salience, and Support for School Bond Referenda.	Survey	Experiment with 4 treatments and control group - vary tax increase framing for bond issue	Random sample of 8,544 US adults
Demaj, L. (2017)	What can performance information do to legislators? A budget-decision experiment with legislators.	Survey	2(value tradeoff) x 2(performance information)	57 legislators
Dineen, J., Robbins, M. D., & Simonsen, B. (2017)	Experimental evidence about deficit reduction strategies: bias in measuring tax and spending preferences.	Survey	2 experiments - both vary tax increase or spending cut	Random sample of US adults: Exp 1 - 355 respondents; Exp 2 - 1,000 respondents
Jilke, S., Lu, J., Xu, C., & Shinohara, S. (2019)	Using Large-Scale Social Media Experiments in Public Administration: Assessing Charitable Consequences of Government Funding of Nonprofits.	Field	Foodbank advertisement: Control ad (no funding info), placebo (donation supported), and treatment (government supported)	296,121 Facebook users in 600 clusters
Kriz, K (2014)	Anchoring and Adjustment Biases and Local Government Referenda Language.	Survey	3 treatments varies the way that cost information is presented for referenda	52 students
Kriz, K. & Clark C.B. (2018)	Does Affect Have an Effect? Willingness to Pay and Psychophysical Numbing.	Survey	2 budget experiments: 2(affect) x 2 (quantiative information)	1,008 MTURK & 1009 MTURK respondents

Leland, S., Piatak, J., & Mohr, Z. (2019)	Accountability in Government Contracting Arrangements: Experimental Analysis of Blame Attribution across Levels of Government.	Survey	2(service provision) x 3 (level of government)	667 MTURK respondents
Lerusse, A., & Van de Walle, S. (2019)	The Importance of Non-Price Criteria in Government Contracting with For-Profit Enterprises.	Survey	Discrete choice experiment (three attributes with two levels and one attribute with three levels)	Civil servants - 166 Belgian; 47 Estonian; 118 Norwegian; 125 German
Nielsen, P. A., & Baekgaard, M. (2013)	Performance information, blame avoidance, and politicians' attitudes to spending and reform: Evidence from an experiment.	Survey	Control group and 3 levels of treatment on school performance	844 random Danish city councilors
Olsen, A. L., Hjorth, F., Harmon, N., & Barfort, S. (2018)	Behavioral Dishonesty in the Public Sector.	Survey	2 dice game experiments	441 students in Denmark; 1,091 residents of 10 other countries (YouGov)
Piatak, J., Mohr, Z., & Leland, S. (2017)	Bureaucratic accountability in third-party governance: Experimental evidence of blame attribution during times of budgetary crisis.	Survey	2(service provision) x 2(budget information)	292 students
Robbins, M. D., Simonsen, B., & Feldman, B. (2004)	The impact of tax price on spending preferences.	Survey	Control and two treatments (avg tax price or specific tax price)	1,200 random West Hartford CT voters
Simonsen, W., & Robbins, M. D. (2000)	The influence of fiscal information on preferences for city services.	Survey	Control (budget exercise); two treatments (avg. tax or no tax)	1,200 random Eugene OR residents

3.1 Budget Tradeoff Experiments

In 1982 UNC Charlotte Professor Ted Arrington[5] and his former student David Jordan published a paper in *Public Administration Review* (PAR) on the importance of including cost in evaluating citizen's budgetary preferences. They used random assignment of the voters in Mecklenburg County North Carolina to show that including cost information changes a person's budgetary preferences for local government services. While it is quite simple, the experiment shows that including the per-person cost of services significantly influences citizen budget priorities in the absence of cost information. The experiment predates the behavioral economics work on tax salience (Chetty, Looney, and Kroft, 2009; Finkelstein, 2009); however, the same behavioral mechanisms are clearly at work making people aware of a cost or tax influences citizens' behavior.

The Arrington and Jordan study informs the largest B-E PBFM body of research: experiments on budgetary tradeoffs. In one of their first experiments, Simonsen and Robbins (2000) show a similar result in the context of a randomized budget exercise that took place in Eugene, Oregon. They show that cost information reduces citizen support for the service. In a significant extension of the earlier papers on cost, Robbins, Simonsen, and Feldman (2004) test two key ideas about support for raising taxes on services. The first is that support for raising taxes to support a service will be influenced by whether it is a low cost or high cost service. They find that citizens do support "reasonable" tax increases for lower cost goods and experience "sticker shock" for higher priced goods, which causes them to reduce their support for tax increases for the service. The other key finding was that having an average tax price did not change citizen tax preferences relative to a specific tax price. This experiment shows a significant increase in the sophistication of the experiment, but it clearly builds off of the previous work of Arrington and Jordan (1982) and Simonsen and Robbins (2000).

In another experiment, Dineen, Robbins, and Simonsen (2017) build off of their earlier work and research in psychology (Hardisty, Johnson, and Weber, 2010) that shows that the inclusion of the word "tax" influences the spending-tax tradeoff that citizens are willing to make. This research shows that issue labeling and framing can significantly influence the amount that they believe should come from taxes. For those that believe that the deficit should be reduced by a combination of tax

5. I note the affiliation and the detail here because the author line of the paper attributes authorship to Thomas S. Arrington. However, the author information shows, and my correspondence confirms, that it was Theodore S. Arrington, the long-time chair of the Department of Political Science at UNCC. It would be hard to imagine a journal, like PAR, these days messing up the name of the author and not publishing a correction. This reflects how far the field has come and indicates how ahead of its time this experiment really was.

increases and spending cuts, when the respondent was asked how much should come from taxes the respondents indicated approximately 23% less than when respondents were asked how much should come from spending cuts. In another recent experiment, Brunner, Robbins, and Simonsen (2018) show that getting information on the cost to the taxpayer for a future bond issuance decreases the probability of voting yes by between 5% and 9%. Similar results by Kriz (2014) show that the presentation of bond referenda influences the likelihood of voting for the referenda. Overall, these results show the importance of cost information and more generally the heuristics and biases involved in making these decisions (Kahneman, 2003; Kriz and Clark, 2018).

3.2 Performance Information and Budgeting

Related to the budget trade-off experiments is research on the importance of performance information[6] on spending preferences. Couching their study in the literature on blame avoidance, Nielsen and Baekgaard (2013) show that high and low performance information causes Danish city councilors to say that they will increase spending, but average performance causes the city councilors to indicate a lower level of spending. High performance is also associated with a significantly lower intention of reforming the service as well. In another study of performance information, Baekgaard (2015) looks at expanding service levels and manipulates both cost and performance information about the expansion to Danish citizens. He shows that performance information matters more when it is presented in conjunction with cost information. The study also shows that performance information is more important to users of the service. Building on all of this work, Baekgaard, Serritzlew, and Blom-Hansen (2016) explore mechanisms of fiscal illusion. Their experiments indicate that lack of attention is the likely cause of fiscal illusion and not lack of information. Demaj (2017) conducts an experiment that shows that performance information changes actual legislators' preferences and notes that there are both good and bad normative aspects to these budgetary changes. Taken as a whole, this work has important implications for both the literature on budgeting and public administration broadly.

3.3 Contracting and Purchasing Experiments

Another body of research that has used experiments extensively is contracting and purchasing. Building upon earlier experimental research in public contracting

6. The impact of performance information on decision making is a large and growing area of research for public administration generally. Here we focus on experiments about the impact of performance information on budgeting and spending decisions. Interested readers in general performance information experiments in public administration should see the review by James and Olsen (2017).

(Marvel, 2015a, 2015b; Marvel and Girth, 2016) which looked at performance perceptions and attributions of blame in contracting, Piatak, Mohr, and Leland (2017) looked at factors that influence blame attribution for service failures when they are either contracted out or provided in house and in cases where fiscal stress is and is not identified. The research shows that governments experience less blame when they contract a service and the suggested level of punishment was less during periods of fiscal stress. In an experiment that extends the previous one, the researchers found that blame attribution also varies by level of government (Leland, Piatak, and Mohr, 2019). Both of these experiments were vignette experiments with simple random assignment into the treatments. In terms of government purchasing, a recent experiment in Belgium, Norway, Estonia, and Germany showed that non-price factors also significantly influence the type of service chosen by local purchasing managers to maximize the utility of the service (Lerusse and Van de Walle, 2019). Using a discrete choice methodology, the research shows how they conducted a marginal willingness to pay for multiple non-price factors.

3.4 Accountability and Transparency

Finally, a number of experiments have been conducted on the area of government accountability. One study on rule following used two experiments and a survey analysis to show that written and consistently applied rules significantly increase rule following (Borry et al., 2018). Piatak and Mohr (2019) show that the gender of the supervisor and the employee also influences rule following in the context of University workers. In another set of studies, Olsen and colleagues (2018) use a dice game to show that dishonesty and corruption may be both a micro and macro level phenomenon. In the experiment in Denmark, they show that rule following is negatively related to public service motivation; however, in another experiment in 10 countries they show that individual level behavioral dishonesty is associated with country level measures of corruption. These experiments show that there are both individual level and macro level influences on accountability and corruption that may be important to study experimentally.

As can be seen in Table 1, the majority of experiments in public administration use vignette or survey experiments. However, there was one field experiment on Facebook related to PBFM that looked at whether government funding crowded out donations to non-profits (Jilke, Lu, Xu, and Shinohara, 2019), but we do not consider this a research area as it is only a single paper. And while there are not many true laboratory experiments in the literature (Tepe and Prokop, 2017), there is a great interest and potential for testing formal theories in public administration using laboratory experiments (Anderson and Edwards, 2015). In spite of these limits to the B-E PBFM literature, this section shows that there are several established

areas for B-E research within the field of PBFM, and the methodology is growing in sophistication with at least one example of a discrete choice experiment and field experiment in the field.

4. EXPERIMENTS FROM OTHER FIELDS RELATED TO PBFM

The review of behavioral-experimental research in PBFM indicates that there are more experiments to draw upon than commonly perceived. The research in fields related to PBFM like economics, political science, psychology, and accounting also addresses many of the same issues (Table 2). Here we provide some examples from these related fields to the areas of taxation, budgeting, contracting, and accountability. We undertake this topical review of the broader B-E literature, which is in contrast to other reviews of the B-E research that focus on theoretical aspects of non-standard decision processes (DellaVigna, 2009; Thaler, 2015) or to topics related to economic theory (McCaffery and Slemrod, 2006), because it directly connects to the existing B-E PBFM literature. Additionally, these are broad areas where more related experiments could be conducted. While the review that is done here is not exhaustive of the B-E literature in these fields and the studies are just examples of the research being done in these areas,[7] it shows many connections that can be drawn from other fields and different types of experimental designs. It also shows the interdisciplinary nature of B-E research and the need to connect with other fields.

7. It is important to note here that all of these fields have their own subfields of behavioral-experimental research that stands in contrast to the nascent field of behavioral PBFM. Behavioral economics has several specialized journals like *Journal of Behavioral and Experimental Economics, Experimental Economics* and many others that publish mostly but not exclusively experimental research. In psychology, the field is called Judgement and Decision Making (JDM) and has similarly focused journals such as *Judgment and Decision Making*, *Organizational Behavior and Human Decision Processes* and others. Accounting has specialized journals such as *Behavioral Research in Accounting*, which is the section journal of Behavior and Organizations Section of the American Accounting Association. Likewise, political science also has specialized journals like *Political Behavior*. All of these fields also regularly publish experimental research in their top field journals as well. Therefore, the studies discussed are examples and are not exhaustive of the research that is related to B-E PBFM.

Table 2. Experiments Discussed in Fields Related to PBFM

Author (Date)	Paper Title	Exp. Type	Experiment/ Quasi-Experiment	Sample/Unit of Analysis
Economics and Taxation Experiments				
Blumenschein, K., et al. (2008)	Eliciting willingness to pay without bias: evidence from a field experiment.	Survey/ Field	2(hypothetical choice)x2(purchase price)	133 undergraduates
Blumenschein, K., et al (1998)	Experimental results on expressed certainty and hypothetical bias in contingent valuation.	Survey/ Field	3(real, hypothetical, and cheap talk)x3($15,$45,$80 medication)	267 diabetes patients
Chetty, R., & Saez, E. (2013)	Teaching the tax code: Earnings responses to an experiment with EITC recipients	Field	control group (standard tax prep. procedure) and treatment group (received personalized information about EITC)	Random 43,000 Earned Income Tax Credit recipients
Hardisty, D. J., Johnson, E. J., & Weber, E. U. (2010)	A dirty word or a dirty world? Attribute framing, political affiliation, and query theory.	Survey	3 experiments: #1 varies tax to reflect cost to society vs. cost to make an activity carbon neutral; #2- two treatments: a cheaper airline ticket or a higher airline ticket cost with a carbon offset; #3 varies the order of listing positive thoughts	#1 - 245 online respondents; #2 - 337 online respondents; # 3 - 316 online respondents
Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1990)	Experimental tests of the endowment effect and the Coase theorem.	Field	5 experiments: (i.e. #1 varied the roles of buyer and seller of tokens, mugs and pens).	Students: sample size ranges from 26-117 students
Kahneman, D., & Tversky, A. (1979)	Prospect theory: An analysis of decision under risk.	Survey	14 experiments (i.e. 50% chance to win 1,000 and 50% to win nothing OR 450 for sure)	Students: sample size ranges from 64-95 students
McCaffery, E. J., & Baron, J. (2006)	Thinking about tax.	Survey	Several experiments using within-subject designs (i.e. varies if question was asked using dollars or percentages, etc.)	About 50-200 online subjects

Sussman, A. B., & Olivola, C. Y. (2011)	Axe the tax: Taxes are disliked more than equivalent costs.	Survey	5 experiments: Exp 1 varies the discount; Exp 2 varies the tax/no tax; Exp 3a varies the investment (tax-exempt vs. taxable), Exp 3b varies the equal bond investments (return with tax/return without tax); Exp 4 varies food cost and tax cost; Exp 5 varies uses of tax dollars (positive list, negative list, no list)	#1 - 191 MTURK respondents, mall shoppers and undergraduates; #2 - 351 MTURK respondents; #3a - 117 MTURK respondents, #3b - 47 MTURK respondents; #4 - 196 MTURK respondents; #5 - 943 MTURK respondents, mall shoppers, etc.
Contracting: Framing, Incentives, and Communication				
Christ, M. H., Sedatole, K. L., & Towry, K. L. (2012)	Sticks and carrots: The effect of contract frame on effort in incomplete contracts.	Lab	2 (contract frame) x 2 (contract implementation) + 1 (baseline) between-subjects experimental design	220 graduate and undergraduate accounting classes randomly assigned principle or agent roles.
Douthit, J. D., Kearney, L. W., & Stevens, D. E. (2012)	Can agent cheap talk mitigate agency problems in the presence of a noisy performance measure? An experimental test in a single-and multi-period setting.	Lab	2 (cheap talk, no cheap talk) x 2 (single-period, multi-period)	120 undergraduate and graduate students
Nichol, J. E. (2018)	The Effects of Contract Framing on Misconduct and Entitlement.	Lab	2 contract framing (bonus/penalty) x 2 misreporting opportunity awareness (before effort/after effort)	99 undergraduate students
Budgeting in Other Disciplines				
Gago-Rodríguez, S., & Naranjo-Gil, D. (2016)	Effects of trust and distrust on effort and budgetary slack: an experiment.	Lab	2(high trust, low trust) x 2(high distrust, low distrust). Randomly assigned role of middle-level manager, upper-level manager, role of observers	160 business managers

Malhotra, N., & Margalit, Y. (2010)	Short-term communication effects or longstanding dispositions? The public's response to the financial crisis of 2008.	Survey	3 experiments - #1 varied blame to Rep or Dem or no blame; #2 varied gain or loss from #3 - varied size of stimulus.	Survey respondents: range from 2068 - 2,748 participants
Ozdemir, S., Johnson, F. R., & Whittington, D. (2016)	Ideology, public goods and welfare valuation: An experiment on allocating government budgets.	Survey	Discrete choice experiment (five programs and five levels)	49 U.S. Adults x 10 choices
Stevens, D. E. (2002)	The effects of reputation and ethics on budgetary slack.	Lab	2 pay schemes x 3 levels of asymmetry. Participants acted as managers.	52 students
Accountability Relevant to Public Financial Reporting				
Chung, J. O., & Hsu, S. H. (2017)	The effect of cognitive moral development on honesty in managerial reporting.	Lab	Submitted cost budgets and completed an exit questionnaire containing the DIT scale	57 undergraduate students assigned role of division managers and submitted a cost budget
Evans, J. H., Hannan, R. L., Krishnan, R. & Moser, D. V. (2001)	Honesty in Managerial Reporting.	Lab	3 experiments that vary the contracts of managers (trust contract, high payoff trust contract, modified trust with a hurdle rate)	Exp 1 - 28 MBA students; Exp 2 - 11 participants; Exp 3 - 28 participants
Schreck, P. (2015)	Honesty in managerial reporting: How competition affects the benefits and costs of lying.	Lab	3 treatments (no competition, economic pressure, rivalry) for 10 periods	60 students

4.1 Taxation: Framing and Salience

Taxation is a large and important area of behavioral research in economics and psychology (Kahneman, 2003; Kahneman and Tversky, 1979; McCaffery and Slemrod, 2006). Framing and loss aversion are the critical elements of prospect theory, which is generally regarded as the beginning of behavioral economics research (Thaler, 2015). Prospect theory, which has been repeatedly tested in many

settings, shows that people are more averse to losses than they value equivalent gains (Antonides, Bolger, and Trip, 2006; Kahneman, Knetsch, and Thaler, 1990). Because of this focus on loss and framing, the area of taxation is a very interesting area for B-E PBFM.

Traditional public finance assumes that people respond to taxes exactly as they would to a price increase (Chetty et al., 2009; Ramsey, 1927). However, recent experimental research showed that making sales tax more visible by including tax inclusive price tags reduced demand by 8% (Chetty et al., 2009). This behavioral response would not be anticipated by traditional economic theory, but behavioral research points to the importance of tax salience as important when determining consumer behavior. Similar results have also been found in diverse tax areas such as the earned-income tax credit (Chetty and Saez, 2013) and in response to E-Z pass tolls (Finkelstein, 2009).

Psychology, marketing and political science also have addressed the issue of tax salience but in different ways from the economics experiments that largely use field experiments. McCaffery and Baron (2006) study isolation effects where people respond quickly to a decision about taxes but ignore other logically relevant information. Hardisty and colleagues (2010) show that framing a tax as a surcharge significantly changes a person's preferences for the level of taxation. As can be readily imagined, the dislike for charges framed as taxes falls more heavily on groups that are traditionally averse to taxation, but some research suggests that having even people from these groups think about the positive uses of these taxes mitigates their aversion to taxation (Sussman and Olivola, 2011).

In short, research in these areas is important and directly relevant to taxation research in PBFM. While recent behavioral economics research focuses on individual level decision making processes like tax salience, it tends to use more robust field experiments to show the ecological validity of the findings. The psychology, marketing, and political science research tends to use more survey or vignette experiments to explore more nuanced theories. This tradeoff between methods is discussed further in the section on implications for future research. In spite of research methods and theories tested, these findings have clear impact on public policy and tax policy that are worthy of further consideration for PBFM.

4.2 Contracting: Framing, Incentives, and Communication

Similar to the research in taxation, the research on contracting in fields like accounting also shows that framing is important. The experiments on contracting show that other issues like incentives and communication are also directly relevant

to contract performance. For example, Christ, Sedatole, and Towry (2012) study how contract framing affects effort. They find that penalty contracts give rise to greater distrust than do bonus contracts. Therefore, when contracts are incomplete, penalty contracts lead to lower effort on tasks not governed by the contract than do bonus contracts. Other issues like incentives and communication are also highly important to contract performance. Nichol (2018) studies the effects of incentive contract framing on misreporting and entitlement. She finds that while penalty contracts can sometimes increase effort relative to bonus contracts, they also encourage greater dishonesty when the effort is not successful. Relatedly, Douthit, Kearney, and Stevens (2012) examine the effects of communication during the contracting process. They find that when agents communicate their intended effort level, agents receive higher wages, provide higher effort, and the principal receives higher profits than the control group. The pre-contract communication, which is non-enforceable and therefore considered "cheap talk" by traditional economic theory, can mitigate the moral hazard problem in agency theory.[8]

4.3 Budgeting: Both Big and Small Issues

Budgeting is a topic that is important in other fields related to budget and finance such as political science and accounting. However, the focus and the experiments that have been conducted focus on slightly different things. In accounting, the experiments on budgeting tend to focus on firm level budgeting concerns such as honesty and budgetary slack; whereas, the political science budgeting experiments tend to focus on more broad public policy issues related to government budgets.

In accounting research, experimental studies have shown the connection between managerial qualities and budgetary slack. For example, Stevens (2002) examines the effects of reputation and ethics on budgetary slack. Stevens finds that reputation and ethical concerns reduced budgetary slack. Budgetary slack is negatively associated with reputation and ethical concerns expressed in the exit questionnaire. Subordinates express lower reputation concerns as information asymmetry regarding productive capability increases, but ethical concerns were not diminished with increases in information asymmetry. Additionally, Gago-Rodríguez and Naranjo-Gil (2016) study the effects of trust and distrust on effort and budgetary slack. They find that the more middle managers trust (distrust) their upper-level managers, the more (less) effort they commit to budgetary proposals.

8. Relative to the B-E PBFM literature, the experiments in accounting were much more likely to be lab-based experiments. Particularly, the use of software z-tree that can track interactions among experiment participants is often used. None of the contracting studies that we reviewed, mostly from accounting, were field experiments as is more commonly done in behavioral economics.

Political science also has a concern with governmental budgeting, and experiments can also be used in this context as well. For example, in one experiment Malhotra and Margalit (2010) show people's preferences for budgets in the context of a stimulus bill following the Great Recession. Their experiment shows that opposition discontinuously shifts as the cost rises. However, there is an inflection point around $1 trillion. Opposition to the bill is unaffected for increases from $800 to $900 billion or from $1 to $1.4 trillion. This research shows the importance of considering mass psychology on budget development.

Interestingly, we found a discrete choice experiment in economics (Ozdemir, Johnson, and Whittington, 2016) that is similar to the budgetary tradeoff experiments done in PBFM. This experiment shows that, while conservatives are not willing to pay more in taxes to increase program outcomes, they are willing to pay more in taxes so that equivalent service levels will not go down. This experiment represents an advance in both the theory and experimental methods for budget tradeoff experiments. Additionally, an older but still very relevant literature traces back to contingent valuation methodology (CVM) that uses field experiments to look at issues like the best method for removing hypothetical bias in willingness to pay questionnaires (Blumenschein, Blomquist, Johannesson, Horn, and Freeman, 2008; Blumenschein, Johannesson, Blomquist, Liljas, and O'Conor, 1998). The research in the budgeting area is the most diverse with all three types of experiments being conducted: survey, lab, and field.

4.4 Accountability Relevant to Public Financial Reporting

As can be imagined, accountability is an important topic for experimental accounting. However, here we focus on a couple of papers that may be directly relevant to public financial reporting. Generally, the papers that we found on accountability in financial reporting were associated with managerial honesty.[9] Schreck (2015) investigates how competition affects the benefits and costs of lying. Schreck finds an individual's willingness to report honestly decreases significantly when rivalry is introduced, even if the economic benefits of lying remain constant. In contrast, economic competition only diminished the salience of honesty preferences of male participants in the experiment. Evans, Hannan, Krishnan, and Moser (2001) find that participants often sacrifice wealth to make honest or partially honest re-

9. There is a very interesting paper that looks at the form of financial reporting that is used, but it does not use an experiment or quasi-experiment as previously discussed. However, the paper is interesting and uses a difference in difference design that attempts to mirror the logic of an experiment (Dorn, Gaebler, and Roesel, 2019).

ports, and they generally do not lie more as the payoff to lying increases. The authors replicated their experiment, except the dollar payoffs were increased by a factor of 5, and found similar results. Chung and Hsu (2017) find a positive and linear relationship between honest managerial reporting and cognitive moral development (CMD), using the defining issues test (Rest, 1989). They further find that the incentives offered to managers interact with the cognitive moral development with the best results obtained when there is a high degree of trust in managers with high moral development. These papers that are not focused specifically on public sector accountability and financial reporting have important implications for financial management but are generally conducted in the lab.

5. COMPARING B-E PBFM EXPERIMENTAL RESEARCH DESIGNS WITH DESIGNS FROM OTHER FIELDS

When we look at Table 1 and Table 2, we can see significant aspects of the research design and the sample of the experiments done in different fields. The biggest distinction between the fields is the type of experiment with economics being much more focused on field experiments and accounting being more likely to use laboratory experiments. Psychology, political science, marketing, and public administration were more likely to use survey experiments. However, these categorizations are not universal as one of the more recent B-E PBFM studies (Jilke, Lu, Xu and Shinohara, 2019) is a field study, and the earlier economics studies were more likely to be survey experiments. While we feel that survey experiments are being used appropriately in B-E PBFM studies, we would like to encourage more use of laboratory and field experiments.

Beyond basic research design, we see similar patterns in the experimental design. The early Arrington and Jordan experiment had only a single factor that it was testing – whether cost information mattered. More recent experiments in PBFM have evolved a little bit, but many are still using basic 2x2 designs. Other fields tend to have slightly more complicated designs and have more experiments per paper. Therefore, we believe that there is room to test more factors within B-E PBFM experiments, and encourage replication of experimental tests as is done broadly in the other fields and done in a few B-E PBFM papers (Baekgaard, Serritzlew, and Blom-Hansen, 2016; Borry et al., 2018; Dineen, Robbins and Simonsen, 2017).

In terms of unit of analysis, B-E PBFM seems to be using quite robust samples with most studies using non-student samples, and sample sizes that range from 57 to over 100,000. The other fields are much more likely to use student samples, but it should be noted that student samples are much more often used in laboratory experiments. The sample sizes range from 49 to 43,000. The sample sizes may also

be larger in B-E PBFM because many of them are more recent, and newer online services like YouGov and MTURK make the collection of online samples much easier than in the past.

In conclusion, B-E PBFM studies seem to be more concerned with sample issues such as representativeness. However, B-E PBFM studies could be improved by using different types of research designs and more complex factors being tested. Many of these issues and design choices may stem from the different epistemic traditions, which we discuss in the next section.

6. IMPLICATIONS FOR BEHAVIORAL-EXPERIMENTAL RESEARCH

The research reviewed both in PBFM and in related fields show that there is a wide range of experimental research available upon which to build B-E PBFM. There is a great potential for building truly interdisciplinary research in many of these areas. However, the research reviewed shows that there are significant challenges in terms of both theory and methods used. The implications of differences in methods and theory are reviewed here as a guide to the emerging field of B-E PBFM.

6.1 Opportunities for B-E Research

One of the great advantages of the rise of behavioral economics has been the resurgence of interest in the experimental method and the ability to causally test the mechanisms of a theory. As was noted in the budget and tax sections, the key issues pointed out by prospect theory are directly relevant to many areas of budget and financial management research and practice. However, the literature that we have reviewed shows that B-E research is much more theoretically varied than simple evaluations of prospect theory transposed to budget and finance contexts. Tax framing, budget discontinuities, and even cheap talk may all affect decisions relevant to budget and finance. These characteristics provide an opportunity to test new and borrowed theories relevant to PBFM. Here we sketch out three areas that we believe are the most likely to connect with other fields and impact PBFM theory and practice.

Public budgeting and budgetary trade-offs – The literature on public budgeting and particularly on budgetary trade-offs is the most developed B-E area of research in PBFM (Arrington and Jordan, 1982; Brunner, Robbins, and Simonsen, 2018; Dineen, Robbins, and Simonsen, 2017; Robbins, Simonsen, and Feldman, 2004; Simonsen and Robbins, 2000). It informs studies of performance management and performance budgeting (Baekgaard, 2015; Baekgaard et al., 2016; Demaj, 2017;

Nielsen and Baekgaard, 2013), which suggests that the experimental literature on budgetary tradeoffs is becoming more influential and recognized. However, in spite of this development that spans nearly forty years, it is still mostly cited by PBFM scholars.

The literature that we reviewed in psychology, accounting, and political science shows that there are significant areas for connection. Particularly interesting are psychology articles that have looked at attribute framing around taxes (Hardisty et al., 2010) and fees (Finkelstein, 2009). It seems that there are more opportunities to do tax framing experiments at both the national level as well as the state and local levels because citizens may have different perceptions of accountability and trust at these levels of government (Leland et al., 2019). Additionally, we believe that studying interventions that may help people overcome their reluctance toward taxes (Sussman and Olivola, 2011) is particularly interesting and important.

Contracting – Another significant area of B-E PBFM research concerns the contracting of public services (Marvel and Girth, 2016; Piatak et al., 2017). While most of this research to date has been simple vignette studies, the use of conjoint or discrete choice experiments (Lerusse and Van de Walle, 2019; Ozdemir et al., 2016) show that the experiments are developing in sophistication. The use of these experiments that can use relatively small samples to generate a significant number of observations on several manipulated variables is likely to be a significant advancement for B-E PBFM.

Theoretically, there is much to be learned from looking at experiments in other areas of contracting. While most of the research is grounded in principal-agent theory, the research that we reviewed showed that incentive framing influences the effort given to a contract (Christ et al., 2012; Nichol, 2018) and that contrary to economic wisdom that pre-contract negotiations is not "cheap talk" because it can signal trust (Douthit et al., 2012). The efforts given to experiments in accounting are highly suggestive for theory development in public sector contracting. A theory that seems to bridge the sector divide in contracting is publicness, where research could be built around a continuum of publicness as suggested by Bozeman and Bretschneider (1994) and that has already been partially tested in a laboratory environment (Brewer and Brewer Jr, 2011). Further theoretical and experimental development in this area around incentive framing, discussion, publicness and blame attribution would be extremely valuable for public and private comparisons of contracting practices or levels of effort given to budget preparation as was conducted by Gago-Rodríguez and Naranjo-Gil (2016).

Taxes – Within the other literatures that we looked at, the effect of taxes on behavior is probably the most developed (Chetty et al., 2009; Chetty and Saez, 2013; Finkelstein, 2009) and grounded squarely in behavioral economic theory (Kahneman and Tversky, 1979; Thaler, 2015). While this is the least developed in the PBFM literature from an experimental perspective, there is a great opportunity to learn and apply this in different tax settings. For example, local governments largely use property tax, which is considered to be one of the most hated taxes because of its salience to homeowners (Cabral and Hoxby, 2012). However, property tax may have one of the most direct applications to services that are valued by citizens, which may mitigate the effects of the tax salience (Sussman and Olivola, 2011). A more robust theory of tax salience and service provision (Ozdemir et al., 2016) grounded in B-E research may open up new avenues for tax research in PBFM.

6.2 Challenges for B-E Research

This review of the behavioral-experimental literatures relevant to PBFM suggests that experiments offer unique challenges for researchers doing experiments in this area. While we personally feel that the benefits outweigh the problems, the potential for conflict with reviewers from different epistemic traditions is likely. Our review of the state of PBFM experiments and the development in other areas suggests that there are two main potential areas for conflict: methods and theory. We review these areas and the potential problems that may be encountered to suggest methodological and theoretical pluralism as the B-E PBFM research develops.

Methods – The review that we conducted shows that methodological preferences and even demands within certain fields may drive research designs. For example, much of the recent research in behavioral economics is now using field experiments (Chetty and Saez, 2013). The reason for this is that experiments are particularly well designed for evaluating causality, but some of the treatments, particularly with lab and vignette experiments, may seem artificial. To be able to truly generalize to practice, it is important to test an experiment in the field where treatments are part of a routine activity for people (for example see Chetty and Saez 2013 on tax filing). Therefore, field experiments can achieve both causal tests and also achieve ecological validity.[10]

While field experiments clearly have more ecological validity, it is important to note that survey and lab experiments can still have internal and even external

10. Similar in the experimental sense to generalizability in statistics, ecological validity refers to how similar the experiment is to real world conditions.

validity. Survey and lab experiments are particularly good for understanding the underlying causal mechanisms. Therefore, it is not surprising that the early experiments on prospect theory (Kahneman and Tversky, 1979) and the endowment effect (Kahneman, Knetsch, and Thaler, 1991) used self-reported survey experiments. As Thaler (2015, p. 37-41) notes, economists do not tend to put much stock in self-reported surveys as they tend to care more about what people do than what they say they are going to do. Thaler notes that while Kahneman and Tversky were not immediately recognized as significant because their methods were unconventional, their work was useful for explaining a variety of behaviors that stand out as contrary to economic theory. Thus, survey experiments were useful for the early work and detecting causal mechanisms that needed to be added to economic theory, like loss aversion, and then in subsequent tests they showed the theoretical mechanisms could also be tested in ways that had further ecological validity. This suggests that in a developing field like B-E PBFM that survey experiments are appropriate for developing and testing theories that may one day be further tested with either laboratory or field experiments.

Related to the issue of self-reported experiments and ecological validity is the issue of incentives in experiments. At a pragmatic level, experiments can cost a significant amount of money to incentivize people, and incentivizing a representative sample can be extremely costly. This may be why we have not seen many experiments in PBFM and very few that come from a completely representative sample (for exception see Brunner et al., 2018). In spite of this, it is possible to use weak or no monetary incentives, such as using students in a class, to test the causal mechanism being explored in an experiment. At a theoretical level, the concern may be that if people are not paid for their time and attention then they have every incentive to not pay attention and simply get through with the experiment. Thaler (2015, pp. 48-49) is instructive about the use of these weak incentives. He notes that economists also dismissed the work of Kahneman and Tversky because they did not use large sums of money. If the bets were larger, then surely people would pay more attention and not make silly mistakes. However, he first notes that most of economic choice revolves around relatively small sums of money, but, even when money was attached to the bets, the problems of inconsistent choice (from a rational theoretical perspective) only increased (Grether and Plott, 1979; Thaler, 2015). From an ecological validity standpoint, any bet that does not involve real money lacks mundane realism. However, as Thaler also notes, it would be difficult to get approval to run experiments where you actually made a person experience a financial loss. This is even more relevant when we think about budget scenarios. It would be difficult or impossible for people to understand, much less feel the incentives, for billion and trillion-dollar budgets. Additionally, people often participate in public activities, like voting, at cost to themselves (Gomez, Hansford, and Krause, 2007) for which they receive little or no reward (Olson, 1965). Therefore,

and in contrast to received wisdom in economics, it may be totally appropriate to not pay,[11] or provide only weak incentives, [12] for participation in a public budgeting study and still have confidence in the validity of the study.

Finally, specific to PBFM is the unit of analysis problem. While behavioral economics and the psychological field of judgement and decision making tends to view the individual as the proper unit of analysis, it can be anticipated in the field of PBFM that there may be other cases where the unit of analysis might be a department or a government decision. The use of quasi-experiments such as geographic discontinuities may be able to show behavioral tripwires in aggregate decision making (an example of this can be seen in Brady and McNulty, 2011). This should be encouraged and not discouraged. Likely, the research from other fields such as sociology, networks, communication and marketing are also likely to be important to both practical and theoretical development of B-E PBFM research.

Theories and Framing – Within different fields, there are different theories and accepted ways of conducting and presenting research. Thaler (2015, p. 37) notes that when he, an economist, began working with Daniel Kahneman, a psychologist, they had to take walks around the hills of Palo Alto to discuss "how members of the other profession think, and what it takes to convince them of some finding." So, while the methods necessary to convince the members of another profession were clearly important, the way that psychologist and economists think are also very different. The theories and models used in these fields are very different. For example, economists tend to favor the use of formal models (Thaler, 2015), but the use of experiments to test formal models has not been published in public administration (Tepe and Prokop, 2017). So, what is the appropriate position for a researcher in PBFM that straddles the worlds of economics and public administration? We suggest that, like the use of survey methodology, the use of more general theories like psychology are appropriate, especially as the field develops. It may be easier to communicate the theory in words and provide a limited test of the theory first, and then a formal model with extensive laboratory testing may follow to more clearly set out the assumptions and limits of the theory. Again, we caution researchers and reviewers not to be too dogmatic about theory, and believe that adding more tools for theory development into our collective toolbox would be a good thing -

11. Dearman and Beard (2009) note that the use of strong incentives like payments may induce people to strategically misrepresent their behavior. They also argue that studies that induce misrepresentation should not be exempt from institutional review board approval and should be required to include desensitization procedures for debriefing participants.

12. Thaler (2015) notes that the use of raffles, such as raffling off a good bottle of Bordeaux (i.e. Kahneman et al., 1991), where the expected value is very low is also non-rational from an economic standpoint, but is a highly effective recruitment strategy!

even though it might have short term costs as researchers have to become more familiar with theories outside of their research tradition.

7. CONCLUSION

The results of this research show that PBFM has more experimental studies than may be commonly perceived. Behavioral-experimental research that looks at the influence of cost and performance information on budget tradeoffs is especially well established. Other research streams in PBFM that have used experiments include contracting, performance, and accountability. In this review, we chose to cast our net broadly when looking at this research to show the variety of theories and methods being explored in the PBFM field. When we look at other fields, we see that there is significant overlap in budgeting experiments particularly in psychology and political science. Accounting research heavily uses experiments, particularly in the area of contracting and accountability. Economics and psychology also conduct a significant number of experiments on the behavioral responses to taxes.

When looking at the field of B-E PBFM, experiments may be useful for testing the theories of other disciplines in the unique context of PBFM. In spite of this, challenges to methods from different epistemic traditions may be problematic. By reviewing the development in other fields, we suggest that basic experiments, such as survey or vignette studies, continue to be used in B-E PBFM research, but also encourage the use of laboratory and field experiments. The use of theories and frameworks from other fields are also encouraged as the field of B-E PBFM develops. In short, both theoretical and methodological pluralism are encouraged.

Ultimately, the field of B-E PBFM is a multidisciplinary field and we see opportunities for doing experimental research in some of the most pressing and hot topics in public budgeting, accounting, and financial management. For example, research on budgetary engagement with the public is a big and important topic in PBFM that has many aspects that could be tested experimentally. Different forms of financial reporting (Dorn et al., 2019) and popular financial reporting (Yusuf, Jordan, Neill, and Hackbart, 2013) may also lend themselves to experimental tests. The research in public sector contracting seems to focus on issues of publicness and blame (Piatak, Mohr, and Leland, 2017), but the accounting research tends to focus on framing and incentives. Combining these research streams may provide better ways to reduce contract failure. Finally, as Brunner et al. (2018) show there are even experimental ways to evaluate citizen's preferences for debt levels, which should be pursued in subsequent experiments. In sum, the studies reviewed here suggest that there are many areas that have already been studied experimentally in

PBFM and many more that could be studied with the use of experimental research designs.

ACKNOWLEDGEMENTS

We would like to acknowledge the help of the symposium editors and the helpful comments of the two anonymous reviewers.

REFERENCES

Anderson, D. and B. Edwards (2015) "Unfulfilled Promise: Laboratory Experiments in Public Management Research", *Public Management Review,* 17(10):1518-1542.

Antonides, G., F. Bolger, and G. Trip (2006) "Classroom experiments in behavioral economics", in M. Altman (ed) *Handbook of Contemporary Behavioral Economics: Foundations and Developments*, New York: ME Sharp, pp. 379-404.

Arrington, T. and D. Jordan (1982) "Willingness to Pay Per Capita Costs as a Measure of Support for Urban Services", *Public Administration Review*, 42(2):168-170.

Baekgaard, M. (2015) "Performance Information and Citizen Service Attitudes: Do Cost Information and Service Use Affect the Relationship?", *International Public Management Journal,* 18(2):228-245.

Baekgaard, M., S. Serritzlew, and J. Blom-Hansen (2016) "Causes of Fiscal Illusion: Lack of Information or Lack of Attention?", *Public Budgeting and Finance,* 36(2):26-44.

Banerjee, A., A. Deaton, and E. Duflo (2004) "Wealth, Health, and Health Services in Rural Rajasthan", *American Economic Review,* 94(2):326-330.

Behn, R. (1985) "Cutback Budgeting", *Journal of Policy Analysis and Management,* 4(2):155-177.

Blumenschein, K., G. Blomquist, M. Johannesson, N. Horn, and P. Freeman (2008) "Eliciting Willingness to Pay without Bias: Evidence from a Field Experiment", *The Economic Journal,* 118(525):114-137.

Blumenschein, K., M. Johannesson, G. Blomquist, B. Liljas, and R. O'Conor (1998) "Experimental Results on Expressed Certainty and Hypothetical Bias in Contingent Valuation", *Southern Economic Journal*, 65(1):169-177.

Borry, E., L. DeHart-Davis, W. Kaufmann, C. Merritt, Z. Mohr, and L. Tummers (2018) "Formalization and Consistency Heighten Organizational Rule Following: Experimental and Survey Evidence", *Public Administration,* 96(2):368-385.

Bouwman, R. and S. Grimmelikhuijsen (2016) "Experimental Public Administration from 1992 to 2014: A Systematic Literature Review and Ways Forward", *International Journal of Public Sector Management,* 29(2):110-131.

Bozeman, B. and S. Bretschneider (1994) "The "Publicness Puzzle" in Organization Theory: A Test of Alternative Explanations of Differences between Public and Private Organizations", *Journal of Public Administration Research and Theory,* 4(2):197-224.

Brady, H. and J. McNulty (2011) "Turning Out to Vote: The Costs of Finding and Getting to the Polling Place", *American Political Science Review,* 105(1):115-134.

Brehm, J., S. Gates, and B. Gomez (1998) "Donut Shops, Speed Traps, and Paperwork: Supervision and the Allocation of Time to Bureaucratic Tasks", in G.A Krause and K. Meier (eds) *Politics, policy, and organizations. Frontiers in the scientific study of bureaucracy* Ann Arbor: The University of Michigan Press, pp. 133-159.

Brewer, G. and G. Brewer Jr. (2011) "Parsing Public/Private Differences in Work Motivation and Performance: An Experimental Study", *Journal of Public Administration Research and Theory,* 21(S3):347-362.

Brunner, E., M. Robbins, and B. Simonsen (2018) "Information, Tax Salience, and Support for School Bond Referenda", *Public Budgeting and Finance,* 38(4):52-73.

Cabral, M. and C. Hoxby (2012) "The Hated Property Tax: Salience, Tax Rates, and Tax Revolts", *National Bureau of Economic Research.* (No. w18514).

Chetty, R., A. Looney, and K. Kroft (2009) "Salience and Taxation: Theory and Evidence", *American Economic Review,* 99(4):1145-1177.

Chetty, R. and E. Saez. (2013) "Teaching the Tax Code: Earnings Responses to an Experiment with EITC Recipients", *American Economic Journal: Applied Economics,* 5(1):1-31.

Christ, M., K. Sedatole, and K. Towry (2012) "Sticks and Carrots: The Effect of Contract Frame on Effort in Incomplete Contracts", *The Accounting Review,* 87(6):1913-1938.

Chung, J. and S. Hsu (2017) "The Effect of Cognitive Moral Development on Honesty in Managerial Reporting", *Journal of Business Ethics,* 145(3):563-575.

Cyert, R. and J. March (1963) *A Behavioral Theory of the Firm*, Englewood Cliffs, NJ: Prentice-Hall.

Dearman, D. and J. Beard (2009) "Ethical Issues in Accounting and Economics Experimental Research: Inducing Strategic Misrepresentation", *Ethics and Behavior,* 19(1):51-59.

DellaVigna, S. (2009) "Psychology and Economics: Evidence from the Field", *Journal of Economic Literature,* 47(2):315-372.

Demaj, L. (2017) "What Can Performance Information Do to Legislators? A Budget-Decision Experiment with Legislators", *Public Administration Review,* 77(3):366-379.

Dineen, J., M. Robbins, and B. Simonsen (2017) "Experimental Evidence about Deficit Reduction Strategies: Bias in Measuring Tax and Spending Preferences", *Journal of Public Budgeting, Accounting and Financial Management,* 29(1):78-103.

Dorn, F., S. Gaebler, and F. Roesel (2019) "Ineffective Fiscal Rules? The Effect of Public Sector Accounting Standards on Budgets, Efficiency, and Accountability", *Public Choice*, Advance online publication. https://doi.org/10.1007/s11127-019-00755-8

Douthit, J., L. Kearney, and D. Stevens (2012) "Can Agent Cheap Talk Mitigate Agency Problems in the Presence of a Noisy Performance Measure? An Experimental Test in a Single and Multi-period Setting", *Journal of Management Accounting Research,* 24(1):135-158.

Evans, J., R. Hannan, R. Krishnan, and D. Moser (2001) "Honesty in Managerial Reporting", *The Accounting Review,* 76(4):537-559.

Finkelstein, A. (2009) "E-ztax: Tax Salience and Tax Rates", *The Quarterly Journal of Economics,* 124(3):969-1010.

Flink, C. (2018) "Predicting Budgetary Change: The Effect of Performance Gaps", *Journal of Public Administration Research and Theory,* 29(2):227-237.

Frederickson, H., and K. Smith (2003) *The Public Administration Theory Primer* Boulder, CO: Westview Press.

Gago-Rodríguez, S. and D. Naranjo-Gil (2016) "Effects of Trust and Distrust on Effort and Budgetary Slack: An Experiment", *Management Decision,* 54(8):1908-1928.

Gomez, B., T. Hansford, and G. Krause (2007) "The Republicans Should Pray for Rain: Weather, Turnout, and Voting in US Presidential Elections", *Journal of Politics,* 69(3):649-663.

Grether, D. and C. Plott (1979) "Economic Theory of Choice and the Preference Reversal Phenomenon", *The American Economic Review,* 69(4):623-638.

Grimmelikhuijsen, S., S. Jilke, A. Olsen, and L. Tummers (2017) "Behavioral Public Administration: Combining Insights from Public Administration and Psychology", *Public Administration Review,* 77(1):45-56.

Hardisty, D., E. Johnson, and E. Weber (2010) "A Dirty Word or a Dirty World? Attribute Framing, Political Affiliation, and Query Theory", *Psychological Science,* 21(1):86-92.

Ho, A. (2011) "PBB in American Local Governments: It's More than a Management Tool", *Public Administration Review,* 71(3):391-401.

Hyde, S. (2007) "The Observer Effect in International Politics: Evidence from a Natural Experiment", *World politics,* 60(1):37-63.

James, O., S. Jilke, and G. Van Ryzin (2017) "Introduction: experiments in public management research", in G. Van Ryzin, O. James, and S. Jilke (eds) *Experiments in Public Management Research: Challenges and Contributions*, Cambridge: Cambridge University Press, pp. 3-19.

James, O., and A. Olsen (2017) "Citizens and public performance measures: Making sense of performance information", in G. Van Ryzin, O. James, and S.

Jilke (eds) *Experiments in Public Management Research: Challenges and Contributions*, Cambridge: Cambridge University Press, pp. 270-290.

Jilke, S., J. Lu, C. Xu, and S. Shinohara (2019) "Using Large-Scale Social Media Experiments in Public Administration: Assessing Charitable Consequences of Government Funding of Nonprofits", *Journal of Public Administration Research and Theory, 29*(4):627-639.

John, P. (2017) "Changing how government works: The transformative potential of an experimental public management", in G. Van Ryzin, O. James, and S. Jilke (eds) *Experiments in Public Management Research: Challenges and Contributions*, Cambridge: Cambridge University Press, pp. 476-494.

John, P. (2018) *How Far to Nudge?: Assessing Behavioural Public Policy* Cheltenham, UK: Edward Elgar Publishing.

Jordan, M. (2003) "Punctuations and Agendas: A New Look at Local Government Budget Expenditures", *Journal of Policy Analysis and Management,* 22(3):345-360.

Justice, J., and G. Miller (2011) "Accountability and Debt Management: The Case of New York's Metropolitan Transportation Authority", *The American Review of Public Administration,* 41(3):313-328.

Kahneman, D. (2003) "Maps of Bounded Rationality: Psychology for Behavioral Economics", *American Economic Review,* 93(5):1449-1475.

Kahneman, D., J. Knetsch, and R. Thaler (1990) "Experimental Tests of the Endowment Effect and the Coase theorem", *Journal of Political Economy,* 98(6):1325-1348.

Kahneman, D., J. Knetsch, and R. Thaler (1991) "Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias", *Journal of Economic Perspectives,* 5(1):193-206.

Kahneman, D., and A. Tversky (1979) "Prospect Theory: An Analysis of Decision under Risk", *Econometrica: Journal of the Econometric Society*, 27(2):263-291.

Kremer, M. (2003) "Randomized Evaluations of Educational Programs in Developing Countries: Some Lessons", *American Economic Review,* 93(2):102-106.

Kriz, K. (2014) "Anchoring and Adjustment Biases and Local Government Referenda Language", Proceedings. Annual Conference on Taxation and Minutes of the Annual Meeting of the National Tax Association, 107(2014):1-16.

Kriz, K., and C. Clark (2018) *Does Affect Have an Effect? Willingness to Pay and Psychophysical Numbing*. Paper presented at the Public Administration Review (PAR) Symposium on Behavioral Approaches to Bureaucratic Red Tape and Administrative Burden, George Washington University, Washington D.C.

Leland, S., J. Piatak, and Z. Mohr (2019) *Accountability in Government Contracting Arrangements: Experimental Analysis of Blame Attribution across Levels of Government*. Paper presented at the Public Management Research Conference, Chapel Hill, NC.

Lerusse, A. and S. Van de Walle (2019) *The Importance of Non-Price Criteria in Government Contracting with For-Profit Enterprises*. Paper presented at the Public Management Research Conference, Chapel Hill, NC.

Levine, C. (1978) "Organizational Decline and Cutback Management", *Public Administration Review,* 38(4):316-325.

Li, H., and G. Van Ryzin (2017) "A systematic review of experimental studies in public panagement journals", in G. Van Ryzin, O. James, and S. Jilke (eds) *Experiments in Public Management Research: Challenges and Contributions*, Cambridge: Cambridge University Press, pp. 20-36.

Lipsky, M. (1980) *Street-level Bureaucracy: Dilemmas of the Individual in Public Services,* New York: Russell Sage Foundation Publications.

Malhotra, N., and Y. Margalit (2010) "Short-term Communication Effects or Longstanding Dispositions? The Public's Response to the Financial Crisis of 2008", *The Journal of Politics,* 72(3):852-867.

Marvel, J. (2015a) "Public Opinion and Public Sector Performance: Are Individuals' Beliefs about Performance Evidence-based or the Product of Anti–Public Sector Bias?", *International Public Management Journal,* 18(2):209-227.

Marvel, J. (2015b) "Unconscious Bias in Citizens' Evaluations of Public Sector Performance", *Journal of Public Administration Research and Theory,* 26(1):143-158.

Marvel, J. and A. Girth (2016) "Citizen Attributions of Blame in Third-party Governance", *Public Administration Review,* 76(1):96-108.

McCaffery, E. and J. Baron (2006) "Thinking about Tax", *Psychology, public Policy, and Law,* 12(1):106.

McCaffery, E., and J. Slemrod (2006) *Behavioral Public Finance,* New York: Russell Sage Foundation.

McCue, C. (2000) "The Risk-Return Paradox in Local Government Investing", *Public Budgeting and Finance,* 20(3):80-101.

Miller, G. (1991) *Government Financial Management Theory,* New York: Marcel Dekker.

Miller, G. (1993) "Debt Management Networks", *Public Administration Review,* 53(1):50-58.

Mosher, F. (1956) "Research in Public Administration: Some Notes and Suggestions", *Public Administration Review,* 16(3):169-178.

Nichol, J. (2018) "The Effects of Contract Framing on Misconduct and Entitlement", *The Accounting Review,* 94(3):329-344.

Nielsen, P. and M. Baekgaard (2013) "Performance Information, Blame Avoidance, and Politicians' Attitudes to Spending and Reform: Evidence from an Experiment", *Journal of Public Administration Research and Theory,* 25(2):545-569.

Olsen, A., F. Hjorth, N. Harmon, and S. Barfort (2018) "Behavioral Dishonesty in the Public Sector", *Journal of Public Administration Research and Theory,* 29(4):572-590.

Olson, M. (1965) *The Logic of Collective Action,* Cambridge: Harvard University Press.

Ozdemir, S., F. Johnson, and D. Whittington (2016) "Ideology, Public Goods and Welfare Valuation: An Experiment on Allocating Government Budgets", *Journal of Choice Modelling,* 20:61-72.

Piatak, J., and Z. Mohr (2019) "More Gender Bias in Academia? Examining the Influence of Gender and Formalization on Student Worker Rule Following", *Journal of Behavioral Public Administration, 2*(2):1-11.

Piatak, J., Z. Mohr, and S. Leland (2017) "Bureaucratic Accountability in Third-Party Governance: Experimental Evidence of Blame Attribution during Times of Budgetary Crisis", *Public Administration,* 95(4):976-989.

Ramsey, F. (1927) "A Contribution to the Theory of Taxation", *The Economic Journal,* 37(145):47-61.

Rest, J. (1989) *Development in Judging Moral Issues,* Minneapolis: University of Minnesota Press.

Ridley, C. and H. Simon (1937) "Technique of Appraising Standards", *Public Management,* 19(2):46–49.

Robbins, M., B. Simonsen, and B. Feldman (2004) "The Impact of Tax Price on Spending Preferences", *Public Budgeting and Finance,* 24(3):82-97.

Rubin, I. (2008) *Public Budgeting: Policy, Process, and Politics,* New York: ME Sharpe Inc.
Schreck, P. (2015) "Honesty in Managerial Reporting: How Competition Affects the Benefits and Costs of Lying", *Critical Perspectives on Accounting, 27*:177-188.

Simon, H. (1947) *Administrative Behavior: A Study of Decision-making Processes in Adminstrative Organization*, New York: The Free Press.

Simon, H. (1955) "A Behavioral Model of Rational Choice", *The Quarterly Journal of Economics,* 69(1):99-118.

Simon, H. (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.

Simonsen, W. and M. Robbins (2000) "The Influence of Fiscal Information on Preferences for City Services", *The Social Science Journal,* 37(2):195-214.

Stevens, D. (2002) "The Effects of Reputation and Ethics on Budgetary Slack", *Journal of Management Accounting Research,* 14(1):153-171.

Sussman, A. and C. Olivola (2011) "Axe the Tax: Taxes are Disliked More than Equivalent Costs", *Journal of Marketing Research,* 48(SPL), S91-S101.

Tepe, M. and C. Prokop (2017) "Laboratory experiments: Their potential for public management research", in G. Van Ryzin, O. James, and S. Jilke (eds) *Experiments in Public Management Research: Challenges and Contributions*, Cambridge: Cambridge University Press, pp. 139-164.

Thaler, R. (2015) *Misbehaving: The Making of Behavioral Economics,* New York: WW Norton.

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

Thurmaier, K. and K. Willoughby (2014) *Policy and Politics in State Budgeting,* New York: Routledge.

Waldo, D. (1948) *The Administrative State: A Study of the Political Theory of American Public Administration,* New York: Routledge.

Wildavsky, A. (1964) *The Politics of the Budgetary Process,* Boston: Little, Brown and Co.

Yanow, D. (1996) *How Does a Policy Mean?: Interpreting Policy and Organizational Actions,* Washington D.C.: Georgetown University Press.

Yusuf, J., M. Jordan, K. Neill, and M. Hackbart (2013) "For the People: Popular Financial Reporting Practices of Local Governments", *Public Budgeting & Finance,* 33(1):95-113.