Journal of Behavioral
Public Administration
Vol 5(1), pp. 1-25
DOI: 10.30636/jbpa.51.304

Roundtable Article

The (Missing?) Role of Institutions in Behavioral Public Administration–A Roundtable Discourse

Anthony M. Bertelli*, Norma M. Riccucci, Paola Cantarelli***, Maria Cucciniello†, Christian R. Grose††, Peter John†††, Elizabeth Linos♦, Anjali Thomas♦♦, and Martin J. Williams♦♦♦**

Editor's note: In this roundtable, the contributors discuss the role of institutions (or lack thereof) in behavioral public administration (BPA). In a multidisciplinary discourse, the contributors touch on the many tensions that exist between institutional and behavioral perspectives of public administration. This roundtable is intended to spark additional discourse on the role of institutions in how they parameterize behaviors within or how individual behaviors might, in the aggregate, influence the norms and rules that shape institutions. Here at *JBPA*, we encourage further dialogue on the role of institutions in behavioral studies and holding work from a macro-, meso-, and micro-lens accountable to each another (Jilke et al., 2019). The editorial team at *JBPA* is thankful to Herbert Simon Award (Midwest Political Science Association) winners Anthony Bertelli (2020) and Norma Riccucci (2021) for organizing this thoughtful conversation. We hope that the discussion offered in this roundtable will inspire further inquiry from our readers. We encourage thought leaders in the field of public administration and beyond to continue this conversation here at *JBPA*. Therefore, we are announcing a Call for Papers, in response to this roundtable. Contributing papers can take one of several forms: (1) *Research Letters* (of no more than 2,000 words), for instance, might provide replications of existing work in BPA where the replications newly account for institutional embeddedness. (2) *Perspective and Practices* submissions (generally limited to 4,000 words) should be written as thoughtful responses to the discourse below, and (3) *Research Articles* (up to 8,000 words) can be more thoroughly threshed out theoretical conceits about institutions in BPA.

-William G. Resh, Editor-in-Chief

How Models, Theories, and Institutions Can Create a Legacy for Behavioral Public Administration—Anthony M. Bertelli & Norma M. Riccucci

We recently published a piece in *Public Administration Review* reflecting on the state of the behavioral public administration (or BPA) movement in our scholarly field (Bertelli and Riccucci, 2020). Our essay contends that while experiments can enhance our exploration of what works in practice, they cannot inform public management practice without becoming part of an integrated program that includes both non-experimental research and theory building. In this essay, what we seek to explain is how we think BPA might take two important steps. First, how can BPA scholars build models that inform theories of public administration? Second, how can it treat institutions—which, after all, define the context of public administration—seriously? We think that if BPA takes on these challenges, it will have an important legacy in the field of public administration.

*Pennsylvania State University and Institut Barcelona d'Estudis Internacionals, **Rutgers University,
***Scuola Superiore Sant'Anna, †University of Edinburgh, ††University of Southern California,
††† King's College London, ♦University of California, Berkeley, ♦♦ Georgia Institute of Technology,
♦♦♦ University of Oxford

Address correspondence to Anthony Bertelli at (amb12@psu.edu)

Theory

BPA is explicit in inviting theories from social psychology. To wit, Grimmelikhuijsen, Jilke, Olsen and Tummers (2017, 46) see it as "drawing on recent advances in our understanding of the underlying psychology and behavior of individuals and groups." This sounds like a call for what the Sociologist Arthur Stinchcombe (1968, 49) calls "grand causal imageries". That is, the authors envision a research environment in which broad classes of psychological causes yield explanations for many types of phenomena, such as prosocial behavior or an inability to evaluate risks.

Yet on the ground, the focus of studies has been on the empirical consequences of some aspect of these broad theories—for instance, an incapacity for risk judgment leads to inconsistent evaluations of administrative data. What is missing is that these empirical results don't seem to build into a theoretical framework. Our comments for this roundtable, while possibly provocative, are meant to be constructive. They are intended to help BPA become more than just a label for a collection of studies.

Theories and Models

Theories are made up of models (Clarke and Primo 2012). We think BPA would do well to consciously focus on particular models that inform the theories of public administration. Consider the following theory: Accountable administrative institutions interact with accountable behavior by public managers to produce legitimate governance (Bertelli 2021).

Suppose now that the research question for an experiment is "Does managerialism work better in Britain than in Italy?" Suppose further that the answer is expected to be "yes" because Britain has a political system that values accountability, and managerialist public administration that focuses on accountability for outcomes is more legitimate there. Let us say that this claim is tested through a survey experiment that compares mass perceptions of the legitimacy of an outcome-accountable policy in Britain to a corresponding one in Italy, finding that the legitimacy perceptions of respondents in the mass public are higher in Britain than in Italy.

Is this good BPA? We would argue that it isn't because the theory examined is one level too "high". The connection between the question and theory runs through the idea of accountability in political representation and in governance. What is missing is a model that is, in fact, behavioral.

One possible model is the following. When presented with visible, public performance information, voters in the United Kingdom consider it as evaluability information about government, while those in Italy see it as evaluability information about those who perform the work. The direct connection to government for British citizens is what drives the higher legitimacy views, and without that direct connection, Italian voters don't see this as legitimating government policies.

This model and the theory it attends is tied to the essence of public administration as a scholarly field—a field that makes and examines normative arguments with practical significance. Accountability systems have different value across contexts, and this model can reach normative claims about "good" public administration. For instance, if a system values accountability, its public administration ought to be accountable to representatives and to citizens (Bertelli, 2021).

Our recommendations for BPA research are threefold: BPA researchers should seek to construct models that inform theories of public administration. Then, they should observe that some possible models are behavioral and some non-behavioral even though they inform the same theory. We hasten to add that we do not think that the behavioral models should be extracted from the theory, and then tested and published separately from the non-behavioral ones. It is what we learn from our theories in general that improves our scholarly understanding of public administration and that creates knowledge for practice.

Institutional Explanations for BPA

The second thing we believe that BPA must consider, and do so seriously, is the importance of institutions. Consider these two distinct theories (Bertelli, 2021):

1. Where and when values in institutions are correlated with values achieved through behaviors, performance and outcomes are good.
2. Where and when values in institutions are uncorrelated with values achieved through behaviors, performance and outcomes are bad.

The following classes of models can be used for contributing to each theory:

A. Institutions shape behaviors and these in turn shape performance and outcomes;
B. Institutions interact with behaviors to shape performance and outcomes.

Which class of models, A or B, belongs in BPA?

We argue that both do. Think of a class A study in which the same performance metric is understood differently by managers in regional and municipal offices. This, in turn, leads the offices to produce different outputs, which then lead to different outcomes. A behavioral model linking the metric information to perceptions can hardly be divorced from an institutional model that considers, say, the relative authority that municipal and regional officials have for making output determinations.

Now think of a class B study where the regional and municipal offices have differential performance given the same protests about their output decisions by clients. This performance differential ultimately corresponds to an outcome differential, the study shows. The differential outcomes are quite literally the product of institutions (office levels) and behaviors (protests).

There are some things that BPA can do to treat seriously the institutional necessity of public administration research. As the essays in this roundtable by Paola Cantarelli and Maria Cucciniello and Elizabeth Linos observe, one way is to integrate mixed-methods designs into the program. Qualitative work can be very useful to inductively discover models in the exploratory sense that we envisioned (Bertelli and Riccucci 2020). Without a model that informs a theory, the experiment is an exploration of, say, what works and does not work in management practice.

Observational study cannot be ignored for several reasons that the essays by Anjali Thomas, Martin Williams, and Christian Grose make plain. Experimental work tells us different things than non-experimental work, so both ought to be important to public administration researchers. Behavioral models inform different aspects of a theory, but must be considered in connection with other models that inform that theory.

If it takes non-behavioral models, institutions, and non-experimental methods seriously and cohesively, we think that BPA will be good for quite a lot, especially, as Peter John tells us in his essay, if BPA allies with the broad theoretical approach and the moral philosophical concerns that underpin behavioral public policy.

Is the ultimate contribution of BPA really all that behavioral? We think that the BPA-versus-the-others distinction misses the point of public administration research. It is not about an acceptance of particular research methods, but about how we study our field's enduring questions. Too much effort spent distinguishing BPA from the rest of our scholarly enterprise is counterproductive, we think, because it limits what we can achieve as a field.

Towards an Institutional *and* Behavioral Public Administration: How do Institutions Constrain or Exacerbate Behavioral Biases of Administrators? —Christian R. Grose

Behavioral and personal biases shape how public administrators implement policy decisions; and how they engage the citizenry at the street level. I define behavioral biases in public administration as instances where personal attitudes or behavioral discrimination may exist among public administrators. For example, discrimination against identity or racial groups by public officials would be an example of a behavioral bias often considered in the literature on behavioral public administration but less frequently considered in the literature on the role of institutions in shaping administrative outcomes.

There has been an explosion in the field of public administration and in the social sciences examining micro-foundational behavioral decision-making (e.g., Grimmelikhuijsen et al., 2016; Grose, 2014; Grose 2021; Moynihan, 2018; Nathan & White, 2021; Tummers, 2020). This work on behavioral public administration is incredibly important as it reveals that public administrators and other political elites are subject to the same behavioral biases of regular people, yet this behavioral literature does not always seriously engage the role of political institutions – rules, laws, and behavioral equilibria – in constraining or exacerbating these behavioral biases (Bertelli & Riccuci, 2020; Grose, 2014).

On the other hand, macro-research on political institutions often ignores individual-level behavioral biases in administrative decision-making (see Moynihan, 2018 for a more complete review). Scholars of institutions emphasize equilibria in policy implementation under comparative variation in institutional arrangements, but

do not often engage with research demonstrating and empirically documenting that administrators have behavioral biases. This must change.

Scholars of public administration can chart a new course by theorizing about the behavioral model of decision-making within a system of political institutions; and by acknowledging that individual-level biases influence decisions and choices.[1] For instance, public administrators and other political elites exhibit behavioral biases through discrimination against minority identity groups. Further, individual-level attitudes and behaviors of administrators are affected by social psychological and other biases. Formal and informal political institutions incentivize, shape, and constrain administrative behavior, and these institutions also shape these revealed administrative biases and this discrimination. The next research agenda must embrace the role of institutions and behavioral biases in policy implementation among public administrators and other public officials. A research agenda of institutional-behavioral public administration (IBPA) has the potential to invigorate scholarship on administration, institutional design, and behavioral biases and attitudes.

Public administrators and public officials have behavioral biases, just like regular people

The novelty of the research on behavioral public administration moves the field forward by acknowledging these psychological, personal, and behavioral biases of public administrators and other public officials. Public administrators and public officials exhibit racial and ethnic biases (Acolin, Bostic, & Painter, 2016; Butler & Broockman, 2011; Costa, 2017; Dinesen, Dahl, & Schiloer, 2021; Einstein & Glick, 2017; Giulietti, Tonin, & Vlassopoulos, 2015; Jenkins, Landgrave, & Martinez, 2020; Jilke, Van Dooren, & Rys, 2018; Mendez & Grose, 2018; Pfaff et al., 2020; Riccucci & Van Ryzin, 2016; Rodriguez & Rossel, 2018). There is also evidence that the attitudes of public officials are susceptible to social psychological cues (Grose et al., 2021; Jilke & Tummers, 2018), framing (Grose & Peterson, 2020; Sheffer et al., 2018), and other behavioral nudges (Avellaneda, 2013; Jilke, Van de Walle, & Kim, 2015; James, John, & Moseley, 2017; Richardson & John, 2021).

Public administrators are subject to behavioral biases, just like regular people. In the field of political science, for instance, there has been an emphasis on distinguishing between regular people, who have often been found to have cognitive or informational limitations that shape their behavior, and who exhibit racial bias in their attitudes; and public administrators and other public officials who have incentives to make decisions without such behavioral biases.

What the field of behavioral public administration has taught us is that these public officials also face many of the same cognitive or information limitations (Belardinellis, Belle, & Cantarelli, 2021; Belle, Cantarelli, & Belardinelli, 2018) that political scientists have observed with the mass public. Yet scholars of institutions and administration not working in this tradition appropriately interrogate this work as downplaying the role of institutions (Bertelli & Ricucci, 2020). Does it matter if public administrators have cognitive limitations when completing survey experiments that parallel similar behavioral findings among the mass public? Does attitude change for public administrators due to behavioral nudging matter if the institutions in which they implement policy are not also considered?

The attitudes and micro-level behaviors of public administrators are in many ways like those of the mass public. Yet, public officials are strategic when they respond to incentives embedded within their political and administrative institutions. In fact, institutions and principal-agent relationships may, under some circumstances, encourage strategic and rational responses by public administrators (Leaver, 2009). Behavioral and psychological biases that many public administrators have, as shown in novel experimental research, can be constrained by institutions or may be exacerbated by institutions. Institutions bring the importance of behavior that will enhance public administrators' strategic motivations and calculations to the fore; and thus some institutions may serve to reduce these well-established behavioral biases through institutional checks. However, in other contexts, institutional design and the constraints placed on administrators may in fact exacerbate behavioral biases.

In short, institutions and discretion matter for how behavioral biases of public administrators affect their decisions. The field should adopt an institutional-behavioral public administration (IBPA) approach in order to probe the ability for institutions to reduce biases among administrators; and to examine where institutions may exacerbate these well-established attitudinal and behavioral biases among administrators.

Racial bias among public administrators and policy makers

For example, consider racial biases by local election administrators. Experimental audit studies of these administrators suggest they exhibit biases against people of color relative to whites (Hughes et al., 2020; White, Nathan, & Faller, 2015). Local election administrators across the United States have been found to respond less frequently to inquiries from people of color than to white people. This behavior is consistent with other audit studies of discrimination in other domains (e.g., Acolin, Bostic, & Painter, 2016; Butler & Crabtree, 2021; Costa, 2017; Gaddis, 2015; Pfaff et al., 2020; Slough, 2018).

Discrimination among public officials is a significant normative problem that has been empirically and causally identified by scholars. Yet scholars of behavioral public administration have often stopped with diagnosing individual-level racial biases and discrimination. Can this discriminatory behavior be curbed through institutional design, or are these racial and ethnic biases made worse through some administrative and institutional arrangements?

While these randomized audit studies demonstrate discrimination by such officials, what role do institutions and principal-agent relationships play? In instances in which institutions serve to exacerbate discrimination, the behavioral biases of public administrators are even more normatively troubling. But scholars have not frequently linked racial bias by public officials uncovered via behavioral experiments to broader institutions or policy choices (though see Mendez and Grose, 2018). Only rarely have scholars sought to examine ways to curb these discriminatory behaviors uncovered by audit studies (Andersen & Guul, 2019; Butler & Crabtree, 2021; Fang, Guess, & Humphreys, 2019; Landgrave, 2019; Riccucci, 2009).

The same local election administrators who respond less frequently to voters of color in audit studies (White, Nathan, and Faller, 2015) also make administrative decisions on where to open polling locations and on how to implement voter identification and signature-matching laws that could disparately impact voters by race and ethnicity (Atkeson et al., 2010; Suttman-Lea, 2020). Scholars should seek to identify the institutions that curb this revealed discriminatory behavior by administrators; as well as continue causally identifying discriminatory behavior.

Institutions exist that constrain local administrative decisions regarding bias in election administration. For example, prior to 2013, Section 4 and Section 5 of the Voting Rights Act constrained the ability for many local elected officials and local public administrators to make administrative changes (Herd & Moynihan, 2018; Kropf, 2016). Any public administrator or public official in former Section 5-covered jurisdictions seeking to make a change to an election practice, such as the siting or closure of polling locations, had to seek pre-approval from the U.S. Department of Justice or the U.S. District Court for the District of Columbia (Grose, 2011; Middlemass, 2015). Since the closure of polling places may have disparate impacts on African American voters and Latinx voters (Jones, 2020), this constraining institution of federal preclearance likely made it harder for local public administrators to act on their racial biases revealed by behavioral audit studies. The preclearance institution made it harder for local officials to close polling places in communities of color and was associated with more equitable outcomes (Cortina & Rottinghaus, 2019; Davidson & Grofman, 1994; Hajnal, 2009; Jones, 2020). How did this institution of federal preclearance achieve this outcome? It constrained individual-level behavior of local administrators, many of whom had demonstrated previous behavior of racial and ethnic discrimination in their local jurisdictions.

In 2013, the U.S. Supreme Court ended this preclearance institution in Shelby v. Holder. In the absence of this institutional constraint post-Shelby, public administrators and public officials no longer had to seek federal pre-approval of any administrative change. These local administrators had much more leeway to make administrative changes in how elections are run outside the scope of federal institutions. Yet these significant institutional constraints on administrative behavior are divorced from the scholarship on discrimination in individual-level behavioral public administration.

This is just one example of the role of institutions in shaping and sometimes constraining behavioral biases by public administrators. Instead of continuing work at separate tables with behavioral scholars studying individual-level administrative behavior and institutional scholars ignoring individual-level biases, a research agenda on the role of behavioral biases within institutions in public administration must be harnessed. The lack of consideration of the role of institutions in constraining these behavioral choices is an opportunity for scholars interested in understanding how institutions can be designed to improve public administration by reducing racial and other behavioral biases in administrative decisions. This institutional-behavioral public administration

approach to scholarship could examine when and how institutions constrain racial biases in administrative practices.

Another important area of research would also be to examine how institutions and individual-level behavioral biases exacerbate racial discrimination. Again, consider racial biases and the administration of elections and voting rights. Following the 2020 election, several U.S. states proposed or passed restrictive laws on voting (Cui, 2021). These policies barred local election administrators from seeking additional funding to open polling places and banned methods of voting that may make it easier for voters of color to participate in elections. For instance, the Texas legislature in 2021 sought to ban voting methods like drive-through voting used in high-population counties like Harris County, Texas (Houston) with large percentages of voters of color while retaining other voting methods used more frequently by white voters (Gardner, 2021). This policy provision would severely constrain public administrators in large counties from opening new polling places or making other administrative choices that would increase access to voters of color, while simultaneously providing greater opportunity for local election administrators in small, rural jurisdictions to continue offering consistent and higher relative per capita levels of voting access. This policy would potentially increase or maintain discretion in small, rural communities where racial and ethnic biases of local public administrators may be greater; but would effectively limit discretion by local administrators in larger, more racially diverse communities.

This is just one example where political institutions – in this case a U.S. state legislature – could shockingly empower local administrators in whiter or low-population areas to engage in racial discrimination while constraining the actions of other local administrators in larger, racially diverse areas. Scholars have explored the role that a representative bureaucracy plays in providing greater behavioral administrative discretion (e.g., Marvel & Resh, 2015; Meier et al., 2005; Riccucci & Van Ryzen, 2016), but connecting variation in levels of bureaucratic discretion to institutions and laws that affect such behavior differentially is an important avenue for research within a new framework of institutional-behavioral public administration.

By considering both individual-level and institutional factors in administration, scholars can contribute to the academic and public policy debates over how institutions can be designed to reduce discrimination against minority groups, improve representational outcomes for citizens, and thus enhance the quality of public administration.

Does discretion increase behavioral and discriminatory biases in administrative decisions?

One of the most promising areas at the nexus of behavioral public administration and political institutions is to consider the role of administrative discretion. Some institutions constrain administrative behavior substantially, while other institutional arrangements provide opportunities for more discretion. Further, laws and rules constrain or empower street-level bureaucrats to the extent to which they are able to use their own discretion in interpreting these laws and rules.

Scholars of institutions and politics argue that principal-agent relationships constrain administrative choices, yet principals do not have the time nor capacity to monitor all agent decision-making. When discretion for public administrative agents is low due to constraining institutional arrangements or the threat of monitoring or audits, the administrator's personal or behavioral biases may be reduced. In other instances, though, the principal-agent relationship provides significant slack for the agent (Nathan & White, 2021; Bertelli & Grose, 2009; Bertelli & Grose, 2011; Whitford, 2005) and thus greater discretion.

With greater discretion granted to administrative agents comes more normatively troubling opportunities for those public administrators to engage in racially discriminatory or other discriminatory behavior that emanates from their personal, behavioral biases identified in the literature on behavioral public administration. Yet discretion is often considered important for achieving particular policy outcomes and management practices, and has been studied extensively within institutions and administration. Administrators who have high levels of discretion are more likely to stay in their jobs and develop expertise, while those with little discretion will leave government for private sector opportunities (Clinton et al., 2012; Gailmard & Patty, 2007). Intriguingly, though, this same discretion that leads to the development of administrative expertise and retention of experienced administrators also encourages decision-making that is likely prone to behavioral biases (Kogan, 2017). Especially in the area of racism uncovered in audit studies at the micro-individual level, discretion for administrative agents is likely to lead to the ability to act on these racial biases in implementing policy and administering

programs. The role of institutions, politics, and laws in exacerbating or constraining these racial biases should be studied in a new research agenda of institutional-behavioral public administration.

Towards an institutional-behavioral public administration

Scholars should seek to understand how institutions can reduce these behavioral biases, especially those demonstrating discrimination against minority groups by administrators. How can we design institutions to reduce these discriminatory behaviors that may happen in domains of extensive discretion, yet enhance discretion that causes administrators to develop policy and domain expertise?

It is time for scholars to move forward with an institutional-behavioral public administration and develop theories and empirical tests that take both micro-level individual behavioral biases and institutional factors into account. Institutions and their structures can incentivize and exacerbate behavioral outcomes, behavioral biases, and individual-level discriminatory behaviors. We must understand administrative behavior within institutions.

Institutional and Behavioral Approaches in the Political Economy of Development—Anjali Thomas

A key mechanism through which institutions influence aggregate developmental outcomes - such as economic growth, poverty levels, and public service provision - is by shaping the behavior of bureaucrats, elected representatives and citizens. Thus, gaining a fine-grained understanding how institutions shape the behavior of these actors is key to addressing questions about when, why and how institutions affect development, questions that lie at the heart of political economy and development economics. However, the effects of institutions are often difficult to study using experimental approaches which has led studies focusing on behavioral public administration to often fail to incorporate an analysis of institutions (Bertelli & Riccucci 2020). In this essay, I discuss three key ways in which experimental approaches have been used to shed light on how institutions affect the behavior of bureaucrats and discuss the trade-offs involved with each approach. The first two approaches involve the use of field experiments or "natural field experiments" (Belle & Cantarelli 2018) which have been successfully used by a number of scholars to answer important questions about how institutions shape behavior. The third approach involves the use of "natural experiments" which involve researchers leveraging situations in which a treatment of interest is assigned in a random or quasi-random approach.

The first approach that could be used to effectively study the effects of institutions on bureaucrats' behavior involves the experimental manipulation of institutional features. While it is usually difficult to manipulate institutions as a whole, some studies have successfully studied the effects of randomized interventions that directly vary aspects of the institutional environment and then examined the effects either on the behavior of state agents or on outcomes that are a function of these agents' behavior (e.g. Banerjee et. al., 2021; Blair et. al., 2019; Olken, 2007). For example, (Banerjee et. al., 2021) study the impact of four interventions aimed at improving police performance – the freezing of police transfers, a weekly day off and duty rotation system, the assignment of community observers, and in-service training. These interventions were randomly assigned across police stations in the Indian state of Rajasthan. They then measured the effects of these interventions on various aspects of police performance including the likelihood a case complaint made by a decoy victim surveyor is registered, crime victims' perceptions of the police, and public perceptions of the police. Studies utilizing such approaches are able to thus directly assess the impact of experimentally altered institutional features on the behavior of bureaucrats.

A second approach to studying how institutions shape bureaucrat behavior involves the experimental manipulation of bureaucratic interactions with key actors who shape the experiences and incentives of bureaucrats. This approach involves utilizing field experiments to assess the impact of interventions that are aimed at altering the way that key actors such as politicians and citizens interact with bureaucrats – for example, interventions that seek to alter the behavior of politicians towards bureaucrats (e.g. Raffler 2021) or the behavior of citizens towards bureaucrats (e.g. Bjorkman & Svensson 2009; Raffler, Posner &Parkerson 2020; Gaikwad, Nellis &Thomas 2020). With this second approach, the causal mechanism linking institutions to bureaucrat behavior is more indirect than than the first approach above. With this second approach, the intervention is

aimed at altering politicians' or citizens' behavior towards bureaucrats which, in the context of certain institutional features of the bureaucracy and the political environment, is theorized to alter bureaucrats' incentives to perform a set of actions. Thus, with this approach, the inferences that can be drawn about institutions are indirect. If the intervention works as intended, one can – by measuring both intermediate and final outcomes – draw inferences about whether the institutional features in question work as theorized. For example, in an ongoing project with Nikhar Gaikwad and Gareth Nellis, we partnered with local NGOs in Mumbai to design and implement two interventions aimed at empowering the city's slum-dwellers to gain access to a municipal water connection. One intervention provided these slum-dwellers with assistance in demonstrating their eligibility for a municipal water connection thereby reducing the bureaucratic hassle costs that often arise when citizens seek to formalize their access to public services. We hypothesized that if the institutional setting is Weberian and provides career incentives to bureaucrats for carrying out their job duties, then such an intervention should increase the incentives of bureaucrats in the water department to respond to requests for a water connection by citizens who successfully demonstrate their eligibility. Consistent with the Weberian view, we found that this intervention resulted in a significant increase in reported site visits by bureaucrats in the water department which was one of the steps involved in providing a water connection. This result allowed us to infer that bureaucrats in the institutional setting we study do respond to meritocratic career incentives. However, we found that the intervention did not significantly increase the likelihood of households ultimately obtaining a municipal water connection, which in turn sheds light on the possibilities and limits of hassle-cost interventions in generating bureaucratic responsiveness in settings where the bureaucracy is politically captured. While field experiments of this nature do not directly manipulate the institutional context, they can shed important light on how certain institutional contexts shape the responsiveness of bureaucratic behavior to experimentally induced changes in incentives. Still, a limitation of such approaches is that they cannot directly study the effects of changes in institutions.

Field experiments such as the ones described above are not always feasible to carry out. More than that, however, both of the above-mentioned approaches raise concerns about generalizability particularly because of unknown heterogeneity in the institutional treatment effects – in other words, the aspects of the political environment being manipulated may interact with other aspects of the institutional context in unknown ways. The replication of similar experimental interventions in varied institutional contexts that are carefully chosen to maximize variation on theoretically important dimensions is a potential strategy to minimize such concerns.

The third approach that can be employed by researchers seeking to examine the effects of institutions on behavior involves leveraging the quasi-random assignment of institutional features. This third approach involves researchers identifying situations in which one or more institutional characteristics are assigned in a quasi-random fashion – in other words, situations that produce "natural experiments". Although (Bertelli & Riccucci, 2020) join (Deaton, 2010) in cautioning against the use of confusing and invalid instruments such as indicators of colonial legacies, there are some settings – usually in sub-national contexts - in which the assumption of quasi-random assignment is either (a) more defensible at the outset or (b) can be probed and at least partially validated through additional statistical analyses. For example, researchers can usefully exploit the exogenous timing of elections or of leadership turnover to study the impact of these changes on the inner workings of the bureaucracy. Previous studies show evidence of political influence on the bureaucracy by examining the relationship between government turnover on bureaucrats' transfers or promotions (e.g. Iyer & Mani, 2012; Brierley 2021). Meanwhile, (Pierskalla et. al., 2020) leverage the exogenous timing of the introduction of elections in Indonesia to examine the effects on bureaucratic promotions for women and minorities. Moreover, several studies have used regression discontinuity designs to analyze the effect of various features - such as incumbency (Klasnja, 2015, Lehne, Shapiro & Van Eynde, 2018; Coviello & Gagliarducci, 2017), alignment between politicians at different tiers government (Thomas Bohlken, 2018, Brollo & Nannicini, 2012), a legislator's alignment with a ruling party (e.g. Thomas Bohlken, 2021) and the effect of the number of principals supervising local bureaucrats (e.g. Gulzar & Pasquale, 2017) - on the implementation of development or infrastructure programs. Still others have used the random assignment of audits based on population thresholds to examine the effects of such audits on various indicators of corruption (Coviello & Mariniello 2014, Hidalgo, Canello, Lima-de-Oliveira 2016). The assumptions of quasi-randomness underlying such designs can often be probed and at least partially defended through the use of auxiliary tests such as balance tests or placebo tests.

One potential limitation of many studies relying on "natural experiments" in practice is that they often lack direct measures of the behavior of individual public officials. However, these studies are frequently able to use outcome measures that are closely tied to bureaucrats' actions to shed meaningful light on the effect of characteristics of the institutional environment on bureaucrats' performance. For example, in a study on how legislators' alignment with a ruling party affects rural road provision (Thomas Bohlken, 2021), I showed that alignment with the ruling party produced a significant increase in completed road projects in the legislator's constituency even controlling for a variable capturing the amount of funds sanctioned for road projects in the constituency. Since the design of the program meant that the completion of road projects falls solely in the hands of the bureaucracy, this result provides evidence that legislators in parliamentary systems have the incentive and capacity to shape the implementation of infrastructure programs using their influence with party members in the executive branch who in turn can exert control over the bureaucracy. Thus, even in the absence of direct measures of bureaucrats' actions, measures of bureaucratically controlled outcomes can offer a valid and reliable way of capturing key aspects of bureaucrats' behavior.

Concerns about generalizability are also relevant for studies using natural experiments. For example, studies relying on regression discontinuity designs focus on incumbents who win elections by a razor thin margin but these incumbents may be different from those who win by a large margin in many ways including in terms of their effect on bureaucratic behavior. Another obvious concern with such studies using natural experiments is that they do not focus on the effects of institutions that have not been assigned in a quasi-random fashion. While this emphasis is usually not a problem for individual studies, it can leave important gaps in knowledge in the aggregate about the effects of institutions that may be most theoretically important or relevant to policy or real-world concerns. This concern with natural experiments is similar to the one highlighted in (Bertelli &Ricucci, 2020) regarding the problem of selection bias entailed in utilizing field experiments or RCTs where the control group consists of those who would have otherwise participated if they had not been randomly excluded from a given program. (Bertelli & Ricucci, 2020) join (Heckman & Smith, 1995) in emphasizing that such designs do not allow for inferences about "hard" or "atypical" cases for whom participation in a program may not be likely in the first place.

The three approaches discussed above – experimental manipulation of institutional features, experimental manipulation of bureaucratic interactions, and leveraging quasi-random assignment of institutional features - represent innovative ways of combining institutional scholarship with experimental behavioral research to study key questions in the field of political economy of development. The power of these approaches lie largely in their ability to credibly isolate and identify the causes that lead to changes in the behavior of bureaucrats – knowledge which can make significant theoretical and practical contributions. Indeed, studies using each of these three approaches have tested important theoretical frameworks and offered compelling evidence regarding how aspects of the institutional environment shape the behavior of bureaucrats. At the same time, however, these studies are typically constrained by having to focus on aspects of the institutional environment that vary due to random assignment either by the researcher or by exogenous factors. Often, these aspects of the institutional environment that can be manipulated this way are narrow or limited in scope and the variations that are captured may not be the ones that are the most substantively important or interesting from a theoretical or policy perspective. Thus, in order to ensure a continued advancement our collective understanding of key questions in the field of political economy of development, non-experimental approaches must also be harnessed to complement and buttress the knowledge gained from studies using the above-mentioned approaches.

Institutional Theory in Behavioral Public Administration: A Three-Stage Approach

—**Maria Cucciniello & Elizabeth Linos**

As many before us have noted, the foundations of behavioral public administration (BPA) span decades (Simon, 1947[2]; Kahneman & Tversky, 1980). However, modern behavioral science and experimental research in public administration are less than a decade old (Belle & Cantarelli, 2018; Bhanot & Linos, 2020; Grimmelikhuijsen, Jilke, Olsen, & Tummers, 2017). It is understandable, then, that early work in this space focuses on establishing the principles of behavioral science and the methodological tools that underpin it. This does not mean, however,

that behavioral experiments are incompatible with theory-driven research. We argue that deep theoretical contributions can and should be found in at least three stages of the BPA approach: (a) in the design of interventions; (b) in the analysis of results; and (c) in their interpretation. In public administration, bringing theory in will often mean bringing an institutional lens to the design, analysis, and interpretation of results. We outline this argument in more detail below.

Institutions in the design of interventions

At the heart of intervention design is a hypothesis (or a series of hypotheses) on what may shift behaviors. BPA scholars have focused on this question in many domains: What makes someone apply for a government job (Linos, 2017)? How different ways to communicate information may shape citizen decisions to act and engage with government (Porumbescu, Cucciniello & Gil-Garcia 2020)? What type of performance feedback motivates a different behavioral response (de Boer et al., 2018)? Choosing what interventions to test should be explicitly linked to understanding context. For example, how someone responds to supervisor feedback depends on the formal and informal institutions in which they received it. Formal institutions are constitutions, contracts, and form of government (e.g., North 1990, 1991; Lowndes 1996) and in our example may include written rules about the frequency of feedback or how feedback is used for promotion and bonuses. Informal institutions include 'traditions, customs, moral values, religious beliefs, and all other norms of behavior that have passed the test of time' (Pejovich 1999, p. 166) and in this case may refer to cultural norms of giving everyone the same score, or gendered expectations on working overtime or what "good" looks like. Designing interventions with an understanding of formal and informal institutions may include a deep-dive into the context using qualitative interviews or observational studies, and may affect hypotheses around what might work or what might backfire given institutional norms. Designing interventions with a deep understanding of context, and therefore institutions, is all the more important with newer commitments to methodological rigor. In a world where preregistration of experimental design is expected, and in an environment where field experiments become more common, testing an infinite number of atheoretical permutations will no longer be possible. As such, the best BPA scholars will justify their intervention design and ground the trade-offs that they inevitably make, in theory.

Institutions in the analysis of results

Perhaps less obviously, a theory-driven understanding of institutions can also play a significant role in how the results of a behavioral experiment are analyzed. In their critique of behavioral experiments, Bertelli and Riccucci (2020) push against focusing on average effects. They are right to do so. "RCTs inform us about average treatment effects, but not about atypical or "hard" cases. In an experimental design, the 25th percentile of the difference between treated and untreated subjects is not the difference in 25th percentiles, so an RCT alone cannot inform us about such differences (Bertelli & Riccucci, 2020 page 3). Although looking at average treatment effects is an important and critical component of understanding "what works," the next wave of behavioral public administration (and of behavioral science more broadly) can and does consider heterogeneous treatment effects: what works for whom. Choosing whom to look at separately is an implicit understanding of institutional context.

As a simple case, let's consider heterogeneous effects by race and gender. The reason we consider race and gender as likely sub-groups that may respond differently to a given intervention is because of the historical, systemic, and institutional contexts that shape gendered and racialized behavior. We expect that over the next decade, the demands for a more nuanced understanding of heterogeneous treatment effects will become central to BPA scholarship. The difference here between data mining, and thoughtful analysis of heterogeneity is an institutional lens.

Institutions in the interpretation of results

As BPA scholarship matures, so do the questions surrounding interpretation of results. Internal validity and crisp causal identification will continue to be important (we hope), but if we are to shift both policy and practice, external validity and generalizability will become more central to the conversation. We note that the use of survey experiments and of fictitious scenarios to induce experimental manipulation does not allow for automatic generalization of the results to more enduring and naturally occurring variations in these conditions. But they can often be a first step in exploring the underlying mechanisms or contexts in which a behavior may

apply. When we ask "so what" or what a study's policy implications are, what we are essentially asking is "under what broader institutional contexts should I think this observed behavioral relationship applies?"

This is an area where collaborations between scholars primarily trained in experiments and those primarily trained in institutional theory can be particularly fruitful. For example, many behavioral scientists have been working on how governments can increase take-up of face masks or vaccines during the COVID-19 pandemic. While successful interventions are often rooted in behavioral science theory, their generalizability depends on whether the specific behavioral mechanism applies in a given institutional context. Using a "social norm" message to encourage people to get vaccinated is only generalizable to other contexts where the broader norm on vaccinations is positive. In contrast, a lottery used to encourage vaccinations may be most generalizable to contexts where residents need an excuse for getting vaccinated that doesn't break with unwritten rules about their other social identities. Bringing an institutional lens to the question of generalizability should produce empirical predictions that future scholars in BPA can test. As such, we encourage the field to embrace conceptual replications - not necessarily direct replications - of BPA scholarship so that we can better understand what works, for whom, in what contexts.

What does this look like in practice?

So how could existing BPA scholarship be extended using the above framework? We take an institution-driven approach to BPA to two long-standing public administration questions: how citizens interact with government agencies and how to retain public sector workers.

Policies targeting citizen-based coproduction promise to enhance the responsiveness of public service delivery by directly incorporating unique experiences and information accumulated by citizen service users into the processes of designing, delivering and evaluating public services (Osborne et al. 2016). The experimental research on co-production has focused primarily on individual-level behaviors, such as recycling, and on citizen participation in general government and planning functions (Kang and Van Ryzin, 2020). Another way to focus more on co-production could be looking at the role played by institutions to encourage (or discourage) co-production. For example, to help schools do more with less, several policies have attempted to encourage members of the public to engage with schools and school districts to assist with governance and service provision issues (Addi-Raccah and Ainhoren 2009).

In a recent paper, Porumbescu, Cucciniello, Belle and Nasi (2020) consider what schools - a formal institution - can do to encourage more parental engagement. They focused on school improvement plans, which are policies that are created annually to communicate school performance and measures adopted by individual schools and school districts to address issues related to student performance. School improvement plans were used because they not only publicly disclose performance information, but also because they provide opportunities for members of the public to engage with public schools. The authors sustain that when individuals understand policies and why they matter, they are more likely to contribute to the policy's success by, for example, engaging in coproduction initiatives. Futhermore, they bring in a behavioral lens to predict that individuals exposed to positive performance information will understand the policy better than those exposed to negative performance information. However, in this study the authors evaluate intentions to engage in co-production with a hypothetical government. Despite efforts to empirically advance the understanding to measure engaging intentions through a novel real effort measure, it remains unclear whether co-production intentions tested through survey experiments accurately predict actual engagement. To shed light on this issue, field experiments would prove useful in testing the degree to which these findings generalize to real world settings, looking more at the role played by institutions and focusing on how institutions encourage or discourage these forms of engagement.

Another important topic in PA relates to the retention of frontline workers. A recent field experiment by Linos, Ruffini, and Wilcoxen (2020) shows that reducing burnout among this population can significantly reduce turnover six months later. Burnout is characterized by high emotional exhaustion, cynicism towards clients, and low personal accomplishment. Much of the existing research suggests that both the informal and formal contexts at work predict who burns out, even among frontline workers who perform similar public services. For example, those who feel like they have someone to talk to at work – a perception that is deeply linked to informal institutional norms – are much less likely to report burnout. Moreover, people who feel undervalued by society report higher levels of burnout. For dispatchers, feeling valued is not only linked to informal social

hierarchies but also related to formal institutional structures: in some cities, dispatchers are organizationally part of law enforcement and thus are better connected to other first responders. In other cities, dispatch centers are a separate entity, often described as a call center, which impacts the social status of dispatchers. Building on these correlational findings, the experiment involved sending weekly nudges to dispatchers in nine cities aiming at affirming social belonging and perceived social support among peers, while also strengthening the value of their common professional identity. As such, the experiment combined insights from behavioral science with an understanding of institutional context to ultimately reduce burnout and turnover. Although successful in this context, questions remain about how generalizable this approach might be in other public sector contexts. Future research could explore where such an approach is likely to succeed, and where it wouldn't, using deep institutional knowledge about how dispatchers are viewed (and view themselves) in different local governments. While empirically testing those hypotheses should involve replication, a robust understanding of institutional contexts helps make effective conceptual replication possible.

Implications for Behavioral Public Administration Research

So why do Bertelli and Riccucci (2020), amongst others, think this is not happening in BPA research? We argue that this is already happening, although granted there is a paucity of scholarship here. The best BPA scholarship already combines a deep understanding of PA theory and institutions in design, analysis, and interpretation of results. But it is also important to recognize that what we are asking of scholars is really quite difficult.

Running a field experiment may be considered the "gold standard" of evaluation, but collaborative research with real public sector or nonprofit partners is challenging, and requires trade-offs in design. A first trade-off is how many treatment arms to include.

Disentangling theoretically-driven mechanisms involves separating out an often limited sample into many treatment arms that, at some point, may lead to sample sizes that are too small and an under-powered study. We do not want to return to a world where causal inferences are made off tiny samples or correlational relationships. This means that while many scholars would love to test multiple separate theory-driven hypotheses in separate treatment arms in the field, there is often a clear limitation of sample size that would make such analyses less rigorous. Just as we expect high quality qualitative research, we expect high quality experimental research in PA. So scholars may turn to other types of experiments to fill a theoretical gap – lab or online experiments, for example, can be very useful at disentangling behavioral mechanisms that may be conflated in a field experiment.

A second trade-off which is particularly important for researchers who want to do practically relevant work is accepting what is institutionally feasible in real contexts. Collaborations with government partners is not the only way to do field experiments, but we believe that collaborative research can be a very effective way of answering research questions that public sector agencies need answered. Such collaborations allow for more institutional relevance and more generalizability, and are therefore particularly important for scholars in BPA who want to incorporate an institutional lens in their work. Such collaborations, however, are very difficult to launch – often taking years – and involve practical trade-offs in what can and can't be tested, due to financial, political, or other feasibility constraints. In the same way that we acknowledge the difficulties and value the contributions associated with primary data collection, we should acknowledge the challenges and value the potential benefits of PA scholars investing in more collaborative research with real public sector agencies.

We recognize, however, that the "tyranny" of experiments may be limiting the potential for BPA to be useful to the field. Using insights from psychology and economics to complement theories of PA is useful in policy design as well as in understanding why some public sector institutions are more successful than others. For example, behavioral science should help shape policy decisions about when and how frequently benefits should be rolled-out, or what paper-work should be required of eligible beneficiaries. Behavioral science can help explain the underlying mechanisms that show correlations between a more representative bureaucracy and more equitable outcomes, by documenting the underlying mechanisms that link bias in decision-making to service delivery. In both cases, these contributions do not require a separate field experiment but lean on previous research in behavioral science, brought to a new context.

We hope that mixed methods will help to leverage the strengths and attenuate the limitations of single method approaches, i.e. experiments. Mixed methods offer the opportunity to rely on different perspectives and richness of data offered. In particular, the use of mixed methods in PA will enhance the quality of a study,

increasing also the familiarity of scholars with a specific empirical context and making the interpretation of results more accurate and nuanced. BPA scholarship is not, and should not be, atheoretical. The future of BPA lies in stronger collaboration between scholars with advanced training in different methods, using different inquiry paradigms, and adopting different theoretical perspectives. Only then can we advance our understanding of complex phenomena and be most useful to practice.

Integrating Politics and Institutions Into Behavioral Public Administration Scholarship Through Mixed-Methods and Research Syntheses—Paola Cantarelli

Building on the seminal works of Herbert Simon (1947), Daniel Kahneman (2011), and Richard Thaler (2008) on bounded rationality, intuitive judgment and choice, and nudging, respectively, behavioral public administration (BPA) has developed remarkably quickly in the past few years. On the one hand, the rate of growth of behaviorally-inspired studies in our field has been so high that it has paved the way for research syntheses already (e.g., Battaglio et al. 2019; Della Vigna and Linos 2020). On the other hand, relevant questions within BPA are still relatively little explored, if explored at all. In this realm, for example, Bertelli and Riccucci (2020) call for more work that integrates politics and institutions into behavioral studies in public administration, especially going beyond experimental scholarship, which accounted for 56 percent of BPA papers in 2019 (Battaglio et al. 2019). This commentary takes up this call by embracing a methodological standpoint and suggesting two primary ways that can help scholars and practitioners bridge the gap between politics and institutions and public administration studies inspired by the behavioral sciences. Those strategies, which are fully discussed below, are mixed-methods designs and research syntheses.

Mixed-methods for BPA scholarship

A first strategy that potentially holds the promise of integrating politics and institutions into BPA behavioral public administration is planning mixed-methods work in which scholars and practitioners work hand in hand. Although experimental work that employs randomization is the gold standard in estimating the average treatment effect of an intervention of some kind, "to different degrees, all causal relationships are context dependent, so the generalization of experimental effects is always at issue" (Shadish, Cook, and Campbell 2002, p. 5). Indeed, experiments can only test the effect of things that can be manipulated, such as the size of a financial bonus or the number of children in a classroom. To the contrary, nonmanipulable events (such as an earthquake) or attributes (such as civil servants' age or citizens' political attitudes) cannot be studied as causes that scholars vary in the context of an experiment. In this bipartition, politics and institutions taken as a unitary construct very likely qualify as nonmanipulable rather than manipulable elements, though some of their inherent features such as norms and values may be more manipulable. Studying the effects of nonmanipulable causes may be much harder than discovering the impact of manipulative things, "nonetheless, nonmanipulable causes should be studied using whatever means are available and seem useful" (Shadish, Cook, and Campbell 2002, p. 8).

The main contribution that experimental designs can provide towards a better understanding of nonmanipulable causes is through analogue experiments and natural experiment. The former manipulates an agent that is similar to the cause of interest. For example, institutions can be manipulated through some of their constituting attributes, such as their public vs. private nature or their primary mission. Belle and colleagues (2021) recently investigated whether individuals' likelihood of lifting their privacy rights in the face of COVID-19 varies based on the public versus private nature of the institution accessing their personal data and the length of time during which records can be used. The publicness of the institution and the length of protection of privacy rights may be examples of political and institutional features. Along the same lines, Horvath et al. (2020) explored whether citizens' willingness to use a COVID-19 contact tracing app depends on such elements as, for instance, which public institution is responsible for maintaining records and the duration of data storage. The use of analogue experiments to account for politics and institutions into BPA seems especially promising in light, for instance, of the extensive literature on work motivation in mission-driven organizations. In fact, much in the same way as institutions are hard to manipulate, work motivation also does not lend itself to a

direct manipulation because assigning different typologies of work motivation to civil servants is largely undoable. Nonetheless, experimental and randomized studies with civil servants are abundant in public administration. This scholarship manipulates motivational antecedents rather than motivation itself (Belle 2013) and has generated evidence that practitioners understand and use. By contrasting a naturally occurring event with a comparison condition, natural experiments then can also be useful in discovering the effect of nonmanipulable causes. Public sector reforms that vary political and/or institutional environments have spread across the world for a long time now and could thus be exploited through this lens. Compared to randomized, controlled trials, these experimental approaches present the advantage of accounting for the fact that politics and institutions are hard to manipulate.

In addition to these experimental designs, observational and qualitative studies certainly provide a unique contribution toward a fine-grained understanding of the link between political and institutional conditions and a behavioral outcome of interest. As the research designs discussed above gives unique information and their strengths and weaknesses are different, scholars and practitioners should strive to combine them in well-planned mixed-methods designs, whose "intrinsic appeal [...] lies in the opportunity to leverage the strengths and attenuate the limitations of single method approaches, thus reconciling the ardent disputes between nomothetic and idiographic perspectives" (Mele and Belardinelli 2019, p. 334). Mele and Belardinelli (2019) provide a parsimonious taxonomy of mixed-methods designs based on the chronological concatenations of methods in a research design, namely parallel or sequential mixed-methods. The former compares and triangulates findings that reply to the same question, although they are obtained through different methods that are conducted in simultaneous but separate studies. The latter entails two or more consecutive phases, whereby quantitative studies are followed by a qualitative work to pursue explanatory goals or qualitative findings are followed by a quantitative effort to reach explanatory ends. Just as non-exhaustive examples, quantitative phases may include observational work, analogue experiments, or natural experiments. Qualitative phases may feature case studies or ethnographic techniques. Recent applications of mixed-methods designs include the study of public agencies (Honig 2019) and the study of isomorphism in the public sector (Belle et al. 2019). Given their multifaceted inherent elements, our understanding of politics and institutions in the context of public administration may well benefit from mixed-methods approaches. For instance, a quanti-qualitative sequential exploratory design can collect experimental or nonexperimental data on a BPA question of interest and then take advantage of qualitative work to account for institutional elements that may support similarities and/or differences of the quantitative findings.

Research Syntheses for BPA Scholarship

A second strategy to successfully integrate politics and institutions into BPA is through a rigorous engagement in quantitative meta-analyses and narrative systematic reviews, possibly across the boundaries of disciplines. About a decade ago, James L. Perry called public administration scholars "to turn to meta-analysis more routinely as a tool for cumulating and assessing what we know about the research questions we study" (Perry 2012, p. 481). By providing a summary effect size of either experimental or observational evidence, meta-analyses have proved to be the most cost-effective method for quantitative research syntheses on a topic of interest. Narrative reviews, then, are helpful in synthesizing qualitative work because they shed light on recurring and outlier elements in variables that are otherwise hard to measure. Combining meta-analyses and narrative reviews on the role of politics and institutions in improving public administration might also nurture the generation of middle range theories that can meaningfully sustain their integration into behaviorally-driven scholarship. Meta-analysis and narrative reviews are, by definition, descriptive ex-post summaries rather than normative exante expectation. Hence, the theoretical implications that can be derived from research syntheses would naturally meet the requirements that Abner and colleagues (2017) identify to make middle range theories functional to the development of grand theories: enough concreteness to generate testable hypotheses, consistency with reality sustained by the derivation from data rather than pure theorizing, and predisposition to generate results that can be further synthesized. Used with parsimony and without being a cure for all problems, middle range theories and grand theories may give their contributions to a meaningful expansion of BPA current realm.

Behavioural Public Administration Meets Behavioural Public Policy—Peter John

Light-touch interventions informed by behavioural science-commonly known as nudges-have become very popular in recent years (Halpern, 2016; John, 2018). The key idea is of myopic individuals who depart from what they would rationally want to do (Thaler & Sunstein, 2009). They need the beneficent policy-maker to get them to where they-and society-would like them to be as rational choosers. All the policy-maker has to do is to optimise collective utility without apparent coercion. From this starting point, many thousands of behavioural interventions have been commissioned and implemented.

Such shiny objects obscure the many debates that economists and others have had about the drivers of public action and how choice and judgement work at individual and collective levels. In particular, there is a long-running contest within economics about how welfare and human freedom intersect with choices in public policy, which appears prominently in the new field of behavioural public policy (BPP). BPP has roots in economics, but broadens out to environmental sciences, public health, development, social policy, and other cognate fields (Oliver, 2013). The journal Behavioural Public Policy encapsulates and promotes this research agenda. BPP researchers do not offer an uncritical hurrah to nudge as they are alert to the problems in moral philosophy of articulating such a view, no matter how attractive are the policy implications. Some are overtly critical, believing that nudge misunderstands the nature of behaviouralism and its roots in welfare economics (Sugden, 2018). In place of nudge, BPP presents a less confident view of the policy-maker, who pays attention to ethical concerns, such as human freedom, which cannot be brushed under the carpet with appeals to effect sizes and external validity of behavioural interventions.

The broad theoretical approach and concerns with moral philosophy in BPP should be central to the allied field of behavioural public administration (BPA). BPA is about applying psychological research to classic public administration problems (Grimmelikhuijsen et al., 2017). One important recent development concerns the biases of public administrators, who conform to well-known heuristics (Cantarelli et al., 2020; Moseley & Thomann, 2021), and it is possible to deploy nudge-style interventions to help correct these biases (Wittels, 2020). This recent turn brings BPA closer to nudge, with bureaucrats as the object of interest rather than citizens. It also brings the more theoretically advanced and more sceptical literature from BPA into clear view.

To encourage synergy between BPP and BPA, this paper does three things. First, it sets out the origins of behavioural economics and BPP, using the insights from the synthetic work of Adam Oliver (2013, 2017). Second, it reports a case study of the work of Robert Sugden (2018), who shows how the behavioural-informed social science is consistent with concerns in welfare economics about efficiency and freedom. Having conveyed how different BPP is from nudge, the final segment of the paper is a reflection on how to study public administration with the insights of behavioural economics but without bias correction. In the end, the behavioural agenda is more consistent with how public administrators, in partnership with elected officials, are balancers of difficult ethical dilemmas, having to make judgements about what might be in the best interests of society. The recognition of the imperfection of making tough calls in public policy could be a route to a more democratically-inspired form of public administration.

The origins of behavioural economics

Economists have always been interested in puzzles, that is ideas that do not conform to the initial propositions of formal theory and rational action. It was through thinking about the inconsistencies within the theory of individual behaviour that behavioural economics emerged. Of course, the foundational experiments of Kahneman and Tverksy were important in challenging the basic structure of choice (Kahneman et al., 1982). Oliver (2017), however, traces the history of such thinking to an earlier period-even back to Adam Smith with his social foundations of political economy. He focuses on the work of Allais (1953, 1990), whose experiments show how people do not maximise expected value but subjective value, and that people cannot calculate the probability of winning gambles when faced with choices that could maximize gains. Even the questioning of the fundament principle of independence in expected utility theory became conventional, as shown in the work of Ellsberg (1961). Kahneman and Tversky build upon these foundations with their experiments showing how people rely on mental shortcuts or rules of thumb, such as preference for the status quo or avoiding losses. Most people make errors in their attempts to avoid complex cost-benefit calculations. This was part of several

challenges to the basic assumptions in decision theory, such as preference reversal (Lichtenstein & Slovic, 1971), whereby people suddenly change preferences but not for any gain.

Once this history is taken into account, the work of Richard Thaler, one of the key architects of nudge, becomes much more comprehensible as making another important set of incremental adjustments to decision-making models, such as over time-inconsistent preferences (Thaler & Shefrin, 1981). It is important to note that the positive reception to his and other behavioural work came from those working in the heart of mainstream economics, as shown by his regular column in the Journal of Economic Perspectives from 1987 to 1990 called 'Anomalies'. Thaler's use of the terms 'homo economics' and 'homo sapiens' is a rhetorical flourish, whereas in practice economists make little use of this distinction. The inclusive approach of mainstream economics explains why behavioural economics has become so popular within the discipline.

Though Allais' and other studies are theoretically inspired, the link to policy and BPP can easily be made, and is seen in the work of Thaler who straddles both, and may also be seen in the work of George Loewenstein (Loewenstein & Chater, 2017). The work of Adam Oliver also shows this mix of policy and theory concerns in BPP. Rather than accepting an easy transfer of behavioural science to nudge-informed public policies, he argues that the use of behavioural policies might in fact not be liberty enhancing. Oliver distinguishes between nudge, which retains some autonomy, from budge which is where behavioural public policies are used with some coercion, then shove where just coercion is used (Oliver, 2015). In this way, a more critical perspective is bought into understanding the behavioural tools of public policy.

Sugden and individual freedom

Such a mix between conventional economics and behaviouralism appears in the work of Robert Sudgen, an experimental economist, whose papers long predate current fashions (Loomes & Sugden, 1982; Sugden & Weale, 1979). As a libertarian, Sudgen believes that economic institutions provide the means by which individuals can follow their own projects without interference. Individuals may have free and prosperous lives, with the state as the neutral arbiter. This is the central project of neo-classical economics with its roots in the work of Adam Smith and J S. Mill. In this formulation, the freedom and efficiency concerns in economic theory are not threatened by losing their foundations in individual utility maximising with consistent and stable preferences. Sudgen (2004) accepts the basic social drivers behind individual behaviour in his American Economic Review paper, 'The Opportunity Criterion: Consumer Sovereignty Without the Assumption of Coherent Preferences'. He shows that the outcomes of welfare maximisation and freedom are satisfied even when there is preference instability. So long as there is some price sensitivity, with free market entry and profit, markets will clear even without stable preferences and the consistent ranking of the utility maximiser. He writes, 'A market, we might say, is a complex system of money pumps, each of which is operated with the intention of extracting value from us, the consumers. Nevertheless, that system gives us opportunity and responsibility - whether or not our preferences meet the standards of rational choice theory' (1030). Economics can throw away the crutches of the rational-actor model, but still derive the limited state and maximise the freedom of individuals to decide what their preferences are. Economic thinking and justification can operate just as well with a boarder and more human-centred account of human action. This of course gets the argument back to Adam Smith.

It is no surprise then that nudge receives short shrift from Sudgen (2004). Like many, he finds the claims for libertarian paternalism to be bogus. Nudge involves the policy-maker choosing the welfare of the individual and using behavioural techniques to alter personal choices, often surreptitiously. By definition it can't be libertarian. Sugden thinks that policy-makers should leave individuals with their own socially-informed choices intact so they may determine their lifestyles for themselves without the state saying how they do it. Sugden brings together these arguments in his recent book, The Community of Advantage: A Behavioural Economist's Defence of the Market (2018), whose title, like Sugden's others, is very much to the point. Markets operate in a behaviourally-informed world, whereby decentralisation characterises the institutional set up and freedom is maintained and celebrated.

A better way to correct bias?

Sugden is an outlier in the field of BPP. But libertarians are not the only thinkers to realise that behavioural economics does not always lead to nudge units or attempts to correct the bias of individuals, and that there might be other ways to do public policy and public administration. There can be a form of BPP that takes

people from their own starting points and works with them in ways that respect their autonomy and freedom rather than override them as in the nudge framework. Nudge polices are not like a menu from which policy-makers can chose but are ways of starting the conversation with the citizen and working along the grain of their preferences. It then might be possible to do nudge in a more democratic way.

Such concerns lead a number of authors to consider alternatives to nudge that take into account citizens as the starting point and to design institutions and policies that are geared to having a better dialogue with them, such as in the programme of think (John et al., 2019). Given the difficulty of scaling up think (for example, with citizen juries), later work called nudge plus shows that nudges can be delivered to encourage an element of self-reflection, so getting the benefits of nudge but with consent (Banerjee & John, 2021; John & Stoker, 2019). In this way, policy-makers and administrators can engage with citizens in ways that are non-paternalist and closer to the ideas of Sugden and other libertarians without necessarily rolling back or limiting the state. Instead, it is possible to broaden out the roles of bureaucrats and politicians.

There is a new set of criteria which can be added to the toolkit of administrators, which is less instrumental, and more cognisant of the role of their role as mediators between citizens and democratic institutions. In this way, preference shaping and leadership strategies by bureaucrats and politicians are more defensible if they are more respectful of autonomy in the first place. In other words, the practice of modern public administration should be interactive and intermediatory, allowing for a less instrumental and more human-centred nudge agenda to emerge, which is closer to how bureaucrats are as human beings with their own biases too. Such concerns link to the European idea of interactive governance where this mutual responsiveness can be built into institutions and elite culture (Torfing et al., 2012). There are a number of think mechanisms, such as consultation requirements or citizens forums, that can be introduced into public institutions to encourage this approach of using behavioural science alongside democratic institutions.

This last comment leads to another branch of BPA, which is the attempt to use the techniques of behavioural science to understand elite decision-making, with its heuristics and biases, which of course goes back to the classic works in public administration (Simon, 1955). It may be possible to correct biases by external agencies or even citizens acting as accountability seekers. Bias correction does not assume the beneficial paternalist, but it still suffers the same problem as the top-down nudger by assuming that the administrator should be manipulated, that is treating administrators as non-reflective and where correction of a pre-determined error is the key objective. But there could be better reason for taking the subjective and socially-driven behaviour of bureaucrats as the starting point for more reflection in the policy process itself, whereby decision-makers can use their hunches and intuitions in a better way. If error correction is the goal, then all BPA scholars have done is to elevate a rational-actor model of decision-making as the ideal to aspire to just like nudge does. It comes down to treating individuals, whether citizens or those who hold public office, with equal concern, which is core to the liberal ideal. In so doing they can be encouraged to engage with broader democratic ideas rather than to see their own behaviour as box ticking and thence an instrumental form of compliance. It comes down to respecting the judgement of public administrators, rather than treating them as prisoners in the iron cage. Such a concern with judgement returns public administration to themes articulated by an earlier generation of scholars, such as Vickers (1965).

External Validity and Institutions in Behavioral Public Administration—Martin J. Williams

One common criticism of the recent vein of research in behavioral public administration (BPA) has been that its methodological focus on experiments has helped researchers achieve internal validity at the cost of external validity, by focusing research on narrow questions in highly controlled situations that may have limited generalizability (e.g. Van de Walle, 2017; Hassan and Wright, 2019; Bertelli and Riccucci 2020). Showing that a particular "nudge" worked in one context is interesting, so the criticism goes, but understanding whether it will also work elsewhere is the real prize (Cartwright and Hardie, 2014). Following the contours of wider debates in the social sciences about randomized controlled trials and the increasing emphasis on causal inference more broadly, scholars have pointed out that one important step towards this is by finding ways to use experiments to identify mechanisms (rather than just the impacts of a specific intervention) and to embed evaluations in broader theories of behavior that might have wider applicability (Grimmelikhuijsen et al, 2017; Van de Walle,

2017; Hassan and Wright, 2019; Bertelli and Riccucci, 2020). In this short commentary, I take for granted that BPA at its best has this potential to identify mechanisms and link to theory, and focus on the question of how we can learn just how widely "generalizable" these mechanisms are.

In another article on how policymakers and researchers ought to think about the external validity of impact evaluations, I have written that the effects of a policy intervention (or other causal process) are best understood as being produced by an interaction of mechanism and context (Williams, 2020).[3] Mechanism here is defined as the causal process or sequence of logical steps through which a policy is intended to operate, and context is defined broadly as all characteristics of the target population, implementing organization, and geographic, social, and temporal situation in which the study is conducted. While many (perhaps most) of these contextual variables are irrelevant to a given policy's impacts since they do not interact with the policy's mechanism, an intervention can have a different intervention in a new context than in the context in which it was originally studied if (and only if) differences in some contextual variables interact with the policy's mechanism. To understand or predict the external validity of interventions thus requires analysts to identify the features of a context that support or interfere with the policy's causal mechanism, and then to assess how these features differ across contexts.

My main contribution in this short piece is to point out that political and bureaucratic institutions are perhaps the most important aspects of context that interact with interventions (and other causal mechanisms) to produce the behavior and outcomes in which BPA is interested. For instance, political institutions such as the existence and character of democratic competition, the rule of law, and media freedom are likely to moderate (in the econometric sense of interacting with, to amplify or diminish) the responsiveness of service beneficiaries and bureaucrats alike to behavioral science-inspired policy interventions. They may also affect how generalizable psychological mechanisms like risk aversion or social comparison manifest themselves in bureaucratic behavior. One could make similar arguments for the moderating effects of: cross-cutting bureaucratic institutions, such as civil service career structures, tenure protections, independent audit and administrative justice authorities, and red tape; organizational structures, processes, and cultures specific to particular organizations, such as the collective decision-making procedures and checks imposed by bureaucratic hierarchies within which individual bureaucrats operate, the existence and implementation of incentive and monitoring systems, the extent of de jure and de facto discretion and autonomy granted to public servants (as well as bureaucrats' experiences of how these have been protected and policed), and organizational cultures (e.g. of fear and rigidity, or of performance and innovation); and broader social norms and expectations such as trust in government institutions and expectations of discrimination. While the precise institutional features that matter for understanding a given intervention or causal mechanism will of course vary, collectively these institutional variables are likely to be of first-order importance for understanding the generalizability of insights from research in BPA.

To illustrate how bureaucratic context can moderate the operation of behavioral biases, consider Dissanayake's (2020) lab-in-field experiment in the United Kingdom's Department for International Development that examines how behavioral factors influence bureaucrats' analysis of the costs and benefits of two hypothetical investment choices. He finds that an information treatment in which the minister expresses preferences over the choice negatively affects the cognitive performance of lower-ranking bureaucrats, but that this effect disappears among more senior officials, possibly because they are more comfortable challenging ministers and pride themselves on a culture of integrity and political independence. Not only do behavioral biases operate differently within a single organization depending on hierarchical position, then, but one could also imagine that the mitigating effect of seniority might be weaker in other bureaucracies within the UK with different political-bureaucratic dynamics – for example in ministries with greater domestic political salience – and might be reversed in bureaucracies where junior officials have tenure protections while senior officials are politically appointed, as in the federal civil services of Brazil or the United States.

How can BPA researchers study the interactions between institutions and behavioral mechanisms? One direct approach is to utilize natural variation in contexts, by exploring heterogeneity and regularities in the operation of a given behavioral pattern (or the effectiveness of an intervention to influence a behavioral pattern) across different institutional contexts, either within one study or by aggregating evidence across multiple studies through systematic review and meta-analysis. Another approach is to experimentally vary these institutions

themselves within a given context (as well as using observational causal inference methods on natural experiments, discontinuities, etc.). While experimentally varying institutions poses numerous practical and ethical challenges, there are also opportunities – for instance in conducting field experiments in partnership with government institutions on internal organizational structures, processes, and cultures, or using lab experiments or survey vignettes. Broadening the methodological toolkit, in economics and political science formal models are increasingly used in tandem with experimental data to extrapolate findings out of sample, which Van de Walle (2017) suggests may be useful for BPA. While formal modeling is rarely used in public administration and requires extensive training, other approaches to extrapolation exploit in-sample heterogeneity over observed covariates to adjust estimates to other samples (e.g. Gechter, 2016; Kowalski, 2018; Andrews and Oster, 2018) might be easier to integrate into the public administration toolkit. Qualitative and mixed methods can also complement quantitative analysis to identify mechanisms as well as the contextual assumptions on which the operation of these mechanisms depends – which is the focus of the "realist evaluation" approach (Pawson and Tilley, 1997).[4] Finally, Williams (2020) introduces a tool called "mechanism mapping" which uses a policy's theory of change as the foundation for examining differences across contexts in the key variables on which a policy's mechanism relies. The intuition is that each step in a policy's intended theory of change is actually causal hypothesis whose validity depends on the presence or absence of a set of contextual factors, which is something that can be examined empirically. While this tool is primarily aimed towards supporting policymakers' judgments about the applicability to their own context of a policy intervention that worked elsewhere, the same framework can be used by researchers to guide their research design ex ante and their analysis and theory-building ex post.

The greatest challenge for BPA scholarship in engaging with institutions to improve external validity is their inherent complexity. While the methodological demands of modern causal inference with limited sample sizes mean that studies in BPA (like other quantitatively driven fields) can at best aim to identify the impacts of one or two variables while holding all else constant, perhaps with limited sub-group and interaction analysis across a handful of measured contextual variables, there are dozens or hundreds of institutional variables that might interact with a given behavioral mechanism or intervention. While the problem of engaging with a high-dimensional reality with a low-dimensional research toolkit is daunting enough, the magnitude of the challenge is even greater than the simple number of these variables, because these variables interact with each other. Such complementarities between different aspects of structure and management – for example between incentive practices and the intrinsic motivations of employees that recruitment processes select for – are one of the defining features of organizations (Brynjolfsson and Milgrom, 2013). More broadly, the increasing recognition that interactions among various policy choices, organizational structures, client/stakeholder behaviors, and contextual factors are pervasive in complex service delivery settings like the health or education system has seen the growth of "systems approaches" to studying service delivery (e.g. De Savigny and Adam, 2009; Pritchett 2015; various in Mansoor and Williams, 2021). Such approaches make this complexity the object of study rather than attempting to control it away to focus on a ceteris paribus impact estimate for a single variable. How might comparable recognition of the theoretical complexity of the determinants of behavior influence the methodological development of the relatively RCT-driven field of BPA?

While engaging with institutions is thus imperative for BPA to develop and there exist numerous feasible methods to gradually advance our knowledge about how institutions moderate behavioral mechanisms and interventions, the complexity of institutions should also encourage humility about the generalizability of BPA's insights at all but the most abstract levels. This is not a criticism of BPA or experimental approaches in particular, since the same external validity concerns and arguments about institutional complexity could be applied to other substantive or theoretical questions within public administration. It does, however, emphasize the value of work that is more explicitly comparative across contexts and focused on institutions and organizations themselves. Such research has arguably diminished in status within public administration as quantitative methods and the technical demands of modern causal inference methods have focused increasing attention on individuals as levels of analysis. It also illustrates Van de Walle's (2017) and Moynihan's (2018) points that both BPA's theoretical focus on individual psychology and decision-making and its methodological focus on quantitative experimental methods have encouraged it to focus on these narrower questions. This micro-level focus is fine in itself and often inappropriate, but risks crowding out attention to institutional questions that are less directly amenable to this theoretical and methodological toolkit, both within BPA and within public administration

more broadly as BPA gains increasing prominence. It is thus critical for BPA to reconnect to these institutional questions and methods, despite the challenges and limitations imposed by institutions' inherent complexity.

Researchers in BPA should derive added motivation for confronting these challenges from the potential prize: not just better external validity within BPA, but also the potential for BPA to transcend its current focus on individual-level cognition and behavior and truly grow into a body of research that can provide firmer micro-level foundations for our understanding of higher-level structures and processes (e.,g., of organizations, teams, networks [Ali et al 2021]) within public administration. This would help realize Herbert Simon's initial work in seeking to bridge the gap from individual behavior to organizational and institutional processes and performance (Simon, 1947; Bertelli and Riccucci, 2020), as well as moving BPA closer to being a "design science" that is better able to marshall theory and evidence to help policymakers improve the functioning of the institutions and organizations in which they are embedded (Moynihan, 2018; Bhanot and Linos, 2020).

Indeed, this ability to apply theoretical research insights to actual policy decisions in varied institutional contexts constitutes the true test of external validity for behavioral public administration. Whereas social scientists sometimes envision the challenge of external validity as an attempt to demonstrate the universality of a particular finding (usually in vain), understanding the real-world applicability of research findings always requires engagement with the specificities of context. For behavioral public administration, this means investigating how institutions in all their diversity and complexity moderate the operation of behavioral mechanisms and interventions, and using this to build towards an understanding of how these behavioral mechanisms also shape the institutions within which they are embedded.

Notes

1. To be clear, some research in political science, public policy, and public administration has attempted to bridge this gap between the study of macro-institutions and individual-level behavior and/or has embraced experimental methods most common in the study of behavior to examine institutions (e.g., Avellaneda 2013; Grose and Wood 2020; Levy 2017; Ostrom 1986, 1990; Riker 1967; Fiorina and Plott 1978; Wilson 2011), but more is needed.
2. Bertelli and Riccucci (2020) point out that Simon (1947) brought the behavioral revolution to our field, advocating for applying methods of the natural sciences to human behavior. "It was a bellwether of a micro-level, quantitative, and empirical approach that long preceded BPA" (Bertelli and Riccucci, 2020 p. 1). However, they also note that lots of research in public administration has focused on micro-level behavior, failing to capture the politics which permeate the field of public administration, including human behavior as well as institutional culture and values.
3. Williams (2020) also contains a fuller conceptual discussion of external validity and the wide range of perspectives and literature on it, which are beyond the scope and space constraints of this short roundtable piece.
4. While a review of these methods is beyond the scope of this short paper, see Williams (2020) for a fuller discussion and citations.

References

Abner, G. B., Kim, S. Y., & Perry, J. L. (2017). Building evidence for public human resource management: Using middle range theory to link theory and data. *Review of Public Personnel Administration, 37*, 139-159.

Acolin, A., Bostic, R. & Painter, G. (2016). A Field Study of Rental Market Discrimination Across Origins in France. *Journal of Urban Economics, 95*(1), 49-63.

Addi-Raccah, A., & Ainhoren, R. (2009). School governance and teachers' attitudes to parents' involvement in schools. *Teaching and Teacher Education, 25*(6), 805–813.

Ali, Aisha, Javier Fuenzalida, Margarita Gomez, and Martin J. Williams. (2021) "Four Lenses on People Management in the Public Sector: An Evidence Review and Synthesis." *Oxford Review of Economic Policy*.

Allais, M. (1953). Le Comportement de l'Homme Rationnel devant le Risque: Critique des Postulats et Axiomes de l'Ecole Americaine. *Econometrica, 21*(4), 503–546. https://doi.org/10.2307/1907921

Allais, M. (1990). *Criticism of the postulates and axioms of the American School* (p. 139). Cambridge University Press.

Andersen, S. C. & Guul, T.S.. 2019. Reducing Minority Discrimination at the Front Line – Combined Survey and Field Experimental Evidence. *Journal of Public Administration and Theory, 29*(3), 429-44.

Andrews, I., and E. Oster. 2018. "Weighting for External Validity." NBER Working Paper 23826, National Bureau of Economic Research, Cambridge, MA.

Atkeson, L. R., L. A. Bryant, T. E. Hall, K. Saunders, & R. M. Alvarez. 2010. A New Barrier to Participation: Heterogeneous Application of Voter Identification Policies. *Electoral Studies, 29*(1), 66-73.

Avellaneda, C. N. 2013. Mayoral Decision-making: Issue Salience, Decision Context, and Choice Constraint? An Experimental Study with 120 Latin American Mayors. *Journal of Public Administration Research and Theory, 23*(3), 631-61.

Banerjee, A, Chattopadhyay, R, Duflo, E, Keniston, D and Singh, N. (2021)."Improving Police Performance in Rajasthan, India: Experimental Evidence on Incentives, Managerial Autonomy, and Training." *American Economic Journal: Economic Policy, 13* (1): 36-66.

Banerjee, S., & John, P. (2021). Nudge plus: Incorporating reflection into behavioral public policy. *Behavioural Public Policy*, 1–16. https://doi.org/10.1017/bpp.2021.6

Belardinelli, P., Belle, N. & Cantarelli, P. (2021). The Impact of Boudned Subadditivity on Administrative Behavior among Public and Private Workers. *Public Administration*. <https://doi.org/10.1111/padm.12710>.

Belle, N. (2013). Experimental evidence on the relationship between public service motivation and job performance. *Public Administration Review*, 143-153.

Belle, N., & Cantarelli, P. (2018). Randomized experiments and reality of public and nonprofit organizations: Understanding and bridging the gap. *Review of Public Personnel Administration, 38*(4), 494–511.

Belle, N., Belardinelli, P., Cantarelli, P., & Mele, V. (2019). On iron cages and suboptimal choices: An experimental test of the micro-foundations of isomorphism in the public sector. *International Public Management Journal, 22*(2), 373-414.

Belle, N., Cantarelli, P. & Belardinelli, P. (2018). Prospect Theory Goes Public: Experimental Evidence on Cognitive Biases in Public Policy and Management Decisions. *Public Administration Review, 78*(6), 828-40.

Belle, N., Cantarelli, P., & Battaglio, R. P. (2021). To consent, or not to consent? The publicness effect on citizens' willingness to grant access to personal data in the face of a health crisis. *Journal of European Public Policy*, 1-19.

Bertelli, A. M. & C. R. Grose 2009. Secretaries of Pork? A New Theory of Distributive Public Policy. *Journal of Politics, 71*(3), 926-45.

Bertelli, A. M. & C. R. Grose. 2011. The Lengthened Shadow of Another Institution? Ideal Point Estimates for the Executive Branch and Congress. *American Journal of Political Science, 55*(4), 767-81.

Bertelli, A. M., & Riccucci, N. M. (2020). What Is Behavioral Public Administration Good for? *Public Administration Review*. https://doi.org/10.1111/puar.13283.

Bertelli, Anthony M. 2021. *Democracy Administered: How Public Administration Shapes Representative Government*. Cambridge: Cambridge University Press.

Bhanot, Syon P., and Elizabeth Linos. (2020) "Behavioral Public Administration: Past, Present, and Future". *Public Administration Review 80*(1): 168–71.

Björkman, M. and Svensson, J. (2009). Power to the People: Evidence from a Randomized Field Experiment on Community-Based Monitoring in Uganda, *The Quarterly Journal of Economics, 124* (2), Pages 735–769

Blair, R., Karim, S., & Morse, B. (2019). Establishing the Rule of Law in Weak and War-torn States: Evidence from a Field Experiment with the Liberian National Police. *American Political Science Review, 113*(3), 641-657.

Brierley, S. (2021). Combining Patronage and Merit in Public Sector Recruitment. *The Journal of Politics, 83*(1), 182-197

Brollo, F., & Nannicini, T. (2012). Tying Your Enemy's Hands in Close Races: The Politics of Federal Transfers in Brazil. *American Political Science Review, 106*(4), 742-761.

Brynjolfsson, Erik, and Paul Milgrom. (2013). "Complementarity in Organizations." In Gibbons, Robert, and John Roberts, *The Handbook of Organizational Economics*, Oxford: Princeton University Press, 11-55.

Butler, D. M. & C. Crabtree. 2021. Audit Studies in Political Science. In J. Druckman, & D. Green (Eds.), *Handbook of Advances in Experimental Political Science* (pp. 42-55). New York: Cambridge University Press.

Butler, D. M. & D. E. Broockman. 2011. Do Politicians Racially Discriminate Against Constituents? *American Journal of Political Science, 55*(3), 463-77.

Cantarelli, P., Belle, N., & Belardinelli, P. (2020). Behavioral Public HR: Experimental Evidence on Cognitive Biases and Debiasing Interventions. *Review of Public Personnel Administration, 40*(1), 56–81. https://doi.org/10.1177/0734371X18778090

Cartwright, N., and J. Hardie. (2014) *Evidence-Based Policy: A Practical Guide to Doing It Better*. Oxford: Oxford University Press.

Clarke, Kevin A., and David M. Primo. 2012. *A Model Discipline: Political Science and the Logic of Representations*. Oxford: Oxford University Press.

Clinton, J. D., A. Bertelli, C. R. Grose, D. E. Lewis, & D. C. Nixon. 2012. Separated Powers in the United

States: The Ideology of Agencies, Presidents, and Congress. *American Journal of Political Science, 56*(2), 341-54.

Cortina, J. & B. Rottinghaus. 2019. The Quiet Revolution: Convenience Voting, Vote Centers, and Turnout in Texas Elections. Paper Presented to the Election Sciences, Reform, and Administration Conference, University of Pennsylvania. < https://cpb-us-w2.wpmucdn.com/web.sas.upenn.edu/dist/7/538/files/2019/06/Cortina-and-Rottinghaus-ESRA-2019-Paper.pdf>.

Costa, M. 2017. How Responsive are Political Elites? A Meta-analysis of Experiments on Public Officials. *Journal of Experimental Political Science, 4*(3), 241-54.

Coviello, D. & Mariniello, M. (2014). "Publicity requirements in public procurement: Evidence from a regression discontinuity design, *Journal of Public Economics, 109*, Pages 76-100.

Coviello, Decio, & Stefano Gagliarducci. (2017). "Tenure in Office and Public Procurement." *American Economic Journal: Economic Policy, 9* (3): 59-105.

Cui, I. 2021. The Supreme Court's Voting Rights Act Case Could Gut Civil Rights Protections. Then What? Washington Post, March 15.

Davidson, C. & B. Grofman. 1994. *Quiet Revolution in the South*. Princeton: Princeton University Press.

de Boer, Noortje, Jasper Eshuis, and Erik-Hans Klijn (2018). "Does disclosure of performance information influence street-level bureaucrats' enforcement style?." *Public Administration Review 78*(5), 694-704.

De Savigny, D., & Adam, T. 2009. Systems thinking for health systems strengthening (Alliance Flagship report series). Geneva: World Health Organization.

Deaton, A. (2010). Instruments, Randomization, and Learning about Development. *Journal of Economic Literature 48*: 424–55.

DellaVigna, S., & Linos, E. (2020). Rcts to scale: Comprehensive evidence from two nudge units (No. w27594). National Bureau of Economic Research.

Dinesen, P. T., M. Dahl, & M. Schioler. 2021. When Are Legislators Responsive to Ethnic Minorities? *American Political Science Review, 115*(2), 450-66.

Dissanayake, Ranil. 2020. "The Importance of Being Earnest: Incentives, Noise and Hierarchy in Public Sector Decision-Making." Center the for Study of African Economies conference paper, available at: <https://custom.cvent.com/4E741122FD8B4A1B97E483EC8BB51CC4/files/dissanayaketheimportanceofbeingearnest20200129.pdf>

Einstein, K. L. & D. M. Glick. 2017. Does Race Affect Access to Government Services? An Experiment Exploring Street-level Bureaucrats and Access to Public Housing. *American Journal of Political Science, 61*(1),100-16.

Ellsberg, D. (1961). Risk, Ambiguity, and the Savage Axioms. *The Quarterly Journal of Economics, 75*(4), 643–669. https://doi.org/10.2307/1884324

Fang, A. H., A. M. Guess, & M. Humphreys. 2019. Can the Government Deter Discrimination? Evidence from a Randomized Intervention in New York City. *Journal of Politics, 81*(1), 127-41.

Fiorina, M. P. & C. R. Plott. 1978. Committee Decisions Under Majority Rule: An Experimental Study. *American Political Science Review, 72*(2), 575-98.

Gaddis, S. M. 2015. Discrimination in the Credential Society: An Audit Study of Race and College Selectivity in the Labor Market. *Social Forces, 93*(4), 1451-79.

Gailmard, S. & J. W. Patty. 2007. Slackers and Zealots: Civil Service, Policy Discretion, and Bureaucratic Expertise. *American Journal of Political Science, 51*(4), 873-89.

Gardner, A. 2021. How the Texas Voting Bill Would Have Created Hurdles for Voters of Color. Washington Post, May 30. < https://www.washingtonpost.com/politics/2021/05/30/texas-voting-law/>.

Gechter, M. 2016. "Generalizing the Results from Social Experiments: Theory and Evidence from Mexico and India." Mimeo, Pennsylvania State University, University Park, PA.

Giulietti, C., M. Tonin, & M. Vlassopoulos. 2015. Racial Discrimination in Local Public Services: A Field Experiment in the US. <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2681054>.

Grimmelikhuijsen, S., Jilke, S., Olsen, A. L., & Tummers, L. (2017). Behavioral Public Administration: Combining Insights from Public Administration and Psychology. *Public Administration Review, 77*(1), 45–56. https://doi.org/10.1111/puar.12609

Grose, C. R. & A. K. Wood. 2020. Randomized Experiments by Government Institutions and American Political Development. *Public Choice, 185*(3-4), 401-13.

Grose, C. R. & J. C. Peterson. 2020. Economic Interests Cause Elected Officials to Liberalize Their Racial Attitudes. *Political Research Quarterly, 73*(3), 511-25.

Grose, C. R. 2011. *Congress in Black and White: Race and Representation in Washington and at Home*. New York: Cambridge University Press.

Grose, C. R. 2014. Field Experimental Work on Political Institutions. *Annual Review of Political Science, 17*(1), 355-70.

Grose, C. R. 2021. *Experiments, Political Elites, and Political Institutions*. In J. Druckman, & D. Green (Eds.), Handbook of Advances in Experimental Political Science (pp. 149-64). New York: Cambridge University Press.

Grose, C. R., P. Lopez, S. Sadhwani, & A. Yoshinaka. 2021. Social Lobbying. *Journal of Politics*. <https://doi.org/10.1086/714923>.

Hajnal, Z. L. 2009. Who Loses in American Democracy? *American Political Science Review, 103*(1), 37-57.

Halpern, D. (2016). Inside the Nudge Unit: How Small Changes Can Make a Big Difference. Random House.

Harris, AS., Sigman,R, Meyer-Sahling, J, Mikkelsen, K, Schuster, C. (2020). Oiling the bureaucracy? political spending, bureaucrats and the resource curse, World Development, Volume 127.

Hassan, Shahidul, and Bradley E. Wright. (2020) "The Behavioral Public Administration Movement: A Critical Reflection". *Public Administration Review 80*(1): 163–7.

Heckman, J, and Smith, S. (1995). Assessing the Case for Social Experiments. *The Journal of Economic Perspectives. 9*(2): 85–110.

Herd, P. & D. P. Moynihan. 2018 *Administrative Burden: Policymaking by Other Means*. New York: Russell Sage.

Hidalgo FD, Canello J, Lima-de-Oliveira R. (2016) Can Politicians Police Themselves? Natural Experimental Evidence From Brazil's Audit Courts. *Comparative Political Studies. 49*(13):1739-1773.

Honig, D. (2019). Case study design and analysis as a complementary empirical strategy to econometric analysis in the study of public agencies: Deploying mutually supportive mixed methods. *Journal of Public Administration Research and Theory, 29*(2), 299-317.

Horvath, L., Banducci, S., & James, O. (2020). Citizens attitudes to contact tracing apps. *Journal of Experimental Political Science*, 1–13. https://doi.org/10.1017/XPS.2020.30

Hughes, D. A., M. Gell-Redman, C. Crabtree, N. Krishnaswami, D. Rodenberger, & G. Monge. 2020. Persistent Bias Among Local Election Officials. *Journal of Experimental Political Science, 7*(3), 179-87.

James, O., P. John, & A. Moseley. 2017. *Field Experiments in Public Management*. In O. James, S.R. Jilke, G.G. Van Ryzin (Eds.), Experiments in Public Management Research: Challenges and Contributions (pp. 89-116). New York: Cambridge University Press.

Jenkins, N. R., M. Landgrave, & G. E. Martinez. 2020. Do Political Donors Have Greater Access to Government Officials? Evidence from a FOIA Field Experiment with U.S. Municipalities. *Journal of Behavioral Public Administration, 3*(2). https://doi.org/10.30636/jbpa.32.111.

Jilke, S. & L. Tummers. 2018. Which Clients Are Deserving of Help? A Theoretical Model and Experimental Test. *Journal of Public Administration Research and Theory, 28*(2), 226-38.

Jilke, S., S. Van de Walle, & S. Kim. 2015. Generating Usable Knowledge through an Experimental Approach to Public Administration. *Public Administration Review, 76*(1), 69-72.

Jilke, S., W. Van Dooren, & S. Rys. 2018. Discrimination and Administrative Burden in Public Service Markets: Does a Public-Private Difference Exist? *Journal of Public Administration Research and Theory, 28*(3), 423-39.

Jilke, S., Olsen, A. L., Resh, W. G., Siddiki, S. (2019) Microbrook, Mesobrook, Macrobrook. *Perspectives on Public Management and Governance*, *2*(4), 245–253

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

John, P., & Stoker, G. (2019). Rethinking the role of experts and expertise in behavioural public policy. *Policy & Politics, 47*.

John, P., Cotterill, S., Moseley, A., Richardson, L., Smith, G., Stoker, G., & Wales, C. (2019). *Nudge, Nudge, Think, Think: Experimenting with Ways to Change Citizen Behaviour*. Manchester University Press.

Jones, C. 2020. Voting Rights After Shelby County v. Holder: Analysis of Polling Place Closures. Paper presented at the American Political Science Association virtual meeting.

Kahneman, D. (2011). *Thinking, Fast and Slow*. New York: Routledge.

Kahneman, D., & Tversky, A. (1980). Prospect theory. *Econometrica, 12*.

Kahneman, D., Slovic, S. P., Slovic, P., & Tversky, A. (1982). *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge University Press.

Kang, S., & Van Ryzin, G. G. (2021). *Experimental methods for investigating co-production*. In The Palgrave Handbook of Co-Production of Public Services and Outcomes (pp. 639-657). Palgrave Macmillan, Cham.

Klašnja, M. 2015. Corruption and the Incumbency Disadvantage: Theory and Evidence. *The Journal of Politics 77*(4), 928-942.

Kogan, V. 2017. Administrative Centralization and Bureaucratic Responsiveness: Evidence from the Food Stamp Program. *Journal of Public Administration Research and Theory, 27*(4), 629-46.

Kowalski, A. E. 2018. "How to Examine External Validity Within an Experiment". NBER Working Paper 24834, National Bureau of Economic Research, Cambridge, MA.

Kropf, M. E. 2016. *Institutions and the Right to Vote in America*. New York: Palgrave Macmillan.

Landgrave, M. 2019. Can Constituents Nudge Legislators? Evidence from a Bias Reduction Field Experiment. < https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3439017>.

Leaver, C. 2009. Bureaucratic Minimal Squawk Behavior: Theory and Evidence from Regulatory Agencies. *American Economic Review, 99*(3), 572-607.

Levy, M. K. 2017. Panel Assignment in the Federal Court of Appeals. *Cornell Law Review, 103*(1), 65-116.

Lichtenstein, S., & Slovic, P. (1971). Reversals of preference between bids and choices in gambling decisions. *Journal of Experimental Psychology, 89*(1), 46–55. https://doi.org/10.1037/h0031207

Linos, E. (2017). More Than Public Service: A Field Experiment on Job Advertisements and Diversity in

the Police. *Journal of Public Administration Research and Theory, 28*(1), 67–85.

Linos, E., Ruffini, K., & Wilcoxen, S. (2020). Belonging Affirmation Reduces Employee Burnout and Resignations in Front Line Workers. Working Paper. University of California, Berkeley.

Loewenstein, G., & Chater, N. (2017). Putting nudges in perspective. *Behavioural Public Policy, 1*(1), 26–53.

Loomes, G., & Sugden, R. (1982). Regret Theory: An Alternative Theory of Rational Choice Under Uncertainty. *The Economic Journal, 92*(368), 805–824. https://doi.org/10.2307/2232669

Lowndes, V. (1996). Varieties of new institutionalism: A critical appraisal. *Public Administration, 74*, 181–197.

Mansoor, Zahra, and Martin J. Williams. (2021) "Systems Approaches to Public Service Delivery: Lessons from Health, Education, and Infrastructure". Mimeo, April.

Martin, L and Raffler, P. (2021). "Fault Lines: The Effects of Bureaucratic Power on Electoral Accountability." *American Journal of Political Science, 65* (1), 210–224.

Meier, K. J., E. G. Juenke, R. D. Wrinkle, & J.L. Polinard. 2005. Structural Choices and Representational Biases. *American Journal of Political Science, 49*(4), 758-68.

Mele, V., & Belardinelli, P. (2019). Mixed methods in public administration research: Selecting, sequencing, and connecting. *Journal of Public Administration Research and Theory, 29*(2), 334-347.

Mendez, M. S. & C. R. Grose. 2018. Doubling Down: Inequality in Responsiveness and the Policy Preference of Elected Officials. *Legislative Studies Quarterly, 43*(3), 457-91.

Middlemass, K. M. 2015. The Need to Resurrect Section 5 of the Voting Rights Act. *Journal of Civil Rights and Economic Development, 28*(1), 61-105.

Moseley, A., & Thomann, E. (2021). A behavioural model of heuristics and biases in frontline policy implementation. *Policy & Politics, 49*(1), 49–67. https://doi.org/10.1332/030557320X15967973532891

Moynihan, D. 2018. A Great Schism Approaching? Towards a Micro and Macro Public Administration. *Journal of Behavioral Public Administration, 1*(1). <https://doi.org/10.30636/jbpa.11.15>.

Nathan, N. & A. White. 2021. *Experiments on and with Street-level Bureaucrats*. In J. Druckman, & D. Green (Eds.), Handbook of Advances in Experimental Political Science (pp. 509-25). New York: Cambridge University Press.

North, D. C. (1990). *Institutions, institutional change and economic performance*. New York: Cambridge University Press.

North, D. C. (1991). Institutions. *Journal of Economic Perspectives, 5*, 97–112.

Oliver, A. (2013). *Behavioural Public Policy*. Cambridge University Press.

Oliver, A. (2015). Nudging, Shoving, and Budging: Behavioural Economic-informed Policy. *Public Administration, 93*(3), 700–714. https://doi.org/10.1111/padm.12165

Oliver, A. (2017). *The Origins of Behavioural Public Policy*. Cambridge University Press.

Olken, B. (2007). Monitoring Corruption: Evidence from a Field Experiment in Indonesia. *Journal of Political Economy 115*(2): 200-249.

Osborne, S. P., Radnor, Z., & Strokosch, K. (2016). Co-production and the co-creation of value in public services: A suitable case for treatment? *Public Management Review, 18*(5), 639–653.

Ostrom, E. 1986. *An Agenda for the Study of Institutions*. Public Choice, 48(1), 3-25.

Ostrom, E. 1990. *Governing the Commons*. New York: Cambridge University Press.

Pawson, R., and N. Tilley. (1997) *Realistic Evaluation*. London: SAGE Publications.

Pejovich, S. (1999). The effects of the interaction of formal and informal institutions on social stability and economic devel-opment. *Journal of Markets and Morality, 2*, 164–181

Perry, J. L. (2012). How can we improve our science to generate more usable knowledge for public professionals?. *Public Administration Review*, 479-482.

Pfaff, S., C. Crabtree, H.L. Kern, & J.B. Holbein. 2020. Street-level Bureaucrats Discriminate Based on Religion? *Public Administration Review, 81*(2), 244-59.

Pierskalla, J.H., Lauretig, A., Rosenberg, A.S. and Sacks, A. (2020), Democratization and Representative Bureaucracy: An Analysis of Promotion Patterns in Indonesia's Civil Service, 1980–2015. *American Journal of Political Science*. https://doi.org/10.1111/ajps.12536.

Porumbescu, G. A., Cucciniello, M., & Gil-Garcia, J. R. (2020). Accounting for citizens when explaining open government effectiveness. *Government Information Quarterly, 37*(2), [101451]. https://doi.org/10.1016/j.giq.2019.101451

Porumbescu, G. A., Cucciniello, M., Belle, N., & Nasi, G. (2020). Only hearing what they want to hear: Assessing when and why performance information triggers intentions to coproduce. *Public Administration*, 1-14. https://doi.org/10.1111/padm.12697

Pritchett, Lant. (2015) "Creating Education Systems Coherent for Learning Outcomes: Making the Transition from Schooling to Learning." RISE Working Paper 15/005.

Raffler, Pia. (2021). Does Political Oversight of the Bureaucracy Increase Accountability? Field Experimental Evidence from an Electoral Autocracy. Harvard University Working Paper.

Riccucci, N.M. & G.G. Van Ryzin. 2016. Representative Bureaucracy: A Lever to Enhance Social Equity, Coproduction, and Democracy. *Public Administration Review, 77*(1), 21-30.

Riccucci, N.M. 2009. The Pursuit of Social Equity in the Federal Government: A Road Less Traveled? *Public Administration Review, 69*(3), 373-82.

Richardson, L. & P. John. 2021. Co-designing Behavioural Public Policy: Lessons from the Field about How to 'Nudge Plus.' *Evidence and Policy* <https://doi.org/10.1332/174426420X16000979778231>.

Riker, W.H. 1967. Bargaining in a Three Person Game. *American Political Science Review, 61*(3), 342-56.

Rodriguez, R.P. & C. Rossel. 2018. A Field Experiment on Bureaucratic Discretionary Bias Under FOI Laws. *Government Information Quarterly, 35*(3), 418-27.

Shadish, W. R., T. D. Cook, & D. T. Campbell. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Belmont, CA: Wadsworth Cengage Learning.

Sheffer, L., P.J. Loewen, S. Soroka, S. Walgrave, & T. Sheafer. 2018. Nonrepresentative Representatives: An Experimental Study of the Decision Making of Elected Politicians. *American Political Science Review, 112*(2), 302-21.

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics, 69*(1), 99–118.

Simon, Herbert A. (1947). *Administrative Behavior: a Study of Decision-Making Processes in Administrative Organization*(1st ed.). New York: Macmillan.

Slough, T. 2018. Bureaucrats Driving Inequality in Access: Experimental Evidence from Colombia. <http://taraslough.com/assets/pdf/JMP_appendix.pdf>.

Stinchcombe, Arthur L. 1968. *Constructing Social Theories*. Chicago: University of Chicago Press.

Sugden, R. (2004). The Opportunity Criterion: Consumer Sovereignty Without the Assumption of Coherent Preferences. *American Economic Review, 94*(4), 1014–1033. https://doi.org/10.1257/0002828042002714

Sugden, R. (2018). T*he Community of Advantage: A Behavioural Economist's Defence of the Market*. Oxford University Press.

Sugden, R., & Weale, A. (1979). A Contractual Reformulation of Certain Aspects of Welfare Economics. *Economica, 46*(182), 111–123. https://doi.org/10.2307/2553184

Suttman-Lea, M. 2020. Poll Worker Decision Making at the American Ballot Box. *American Politics Research, 48*(6), 714-18.

Thaler, R. H., & Shefrin, H. M. (1981). An Economic Theory of Self-Control. *Journal of Political Economy, 89*(2), 392–406. https://doi.org/10.1086/260971

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

Thaler, R. H., & Sunstein. (2008). *Nudge: Improving Decisions about Health, Wealth, and Happiness*. New Haven, CT: Yale University Press.

Torfing, J., Peters, B. G., Pierre, J., & Sørensen, E. (2012*). Interactive Governance: Advancing the Paradigm.* In Interactive Governance. Oxford University Press. https://oxford.universitypressscholarship.com/view/10.1093/acprof:oso/9780199596751.001.0001/acprof-9780199596751

Tummers, L. 2020. Behavioral Public Administration. Oxford Research Encyclopedia of Politics. <https://doi.org/10.1093/acrefore/9780190228637.013.1443>.

Van de Walle, Steven. (2017) "The Experimental Turn in Public Management: How Methodological Preferences Drive Substantive Choices". In Experiments in Public Management Research: Challenges and Contributions, Oliver James, Sebastian R. Jilke, and Gregg G. Van Ryzin (eds), Cambridge: Cambridge University Press, 461–75.

Vickers, S. G. (1965). The Art of Judgment: A Study of Policy Making. Basic Books.

White, A.R., N.L. Nathan, & J.K. Faller. 2015. What Do I Need to Vote? Bureaucratic Discretion and Discrimination by Local Election Officials. *American Political Science Review, 109*(1), 129-42.

Whitford, A.B. 2005. The Pursuit of Political Control by Multiple Principals. *Journal of Politics, 67*(1), 39-49.

Williams, Martin J. (2020) "External Validity and Policy Adaptation: From Impact Evaluation to Policy Design." *World Bank Research Observer Volume 35, Issue 2*, 158–191, https://doi.org/10.1093/wbro/lky010.

Wilson, R.K. 2011. The Contribution of Behavioral Economics to Political Science. *Annual Review of Political Science, 14*(1), 201-223.

Wittels, A. (2020). The effect of politician-constituent conflict on bureaucratic responsiveness under varying information frames. In SocArXiv (4x8q2; SocArXiv). Center for Open Science. https://ideas.repec.org/p/osf/socarx/4x8q2.html