ANNUAL REVIEWS

Annual Review of Political Science

Causal Inference, Agency, and the Problem of Inherent Endogeneity

Martin J. Williams

Departments of Organizational Studies and Political Science, University of Michigan, Ann Arbor, Michigan, USA; email: martinjw@umich.edu

Keywords

causal inference, agency, endogeneity, external validity, research methods, complexity

Abstract

Researchers often leverage exogenous variation in an independent variable in order to understand its effects, yet endogenous decision processes are central to many of the real-world phenomena we wish to understand. This review explores whether there are situations in which exogenous and endogenous variation in the same independent variable (e.g., a policy, treatment, or other action) may lead to different outcomes. I begin by laying out a conceptual framework for understanding these inherently endogenous causal processes, identifying three types of mechanisms through which they might arise and discussing their application to a range of empirical phenomena, such as institutional reform, community natural resource governance, and interstate conflict. I then suggest that learning about inherently endogenous causal processes requires researchers to place endogenous decision-making at the center of analysis rather than seeking to abstract away from it. I survey a range of methods (both positivist and nonpositivist) for doing so.

INTRODUCTION

Causal inference research designs often seek to leverage exogenous variation in an independent variable—an action, policy, intervention, or treatment—in order to identify its effect on a dependent variable. This allows researchers to estimate unbiased causal effects by removing the determination of the independent variable from the control of actors in the system being studied, ruling out concerns that actors might endogenously choose actions based on unobserved and thus uncontrolled variables.

At the same time, endogenous choice processes are central to many important real-world phenomena of interest to political scientists. Similarly, one of the main reasons why researchers want to retrospectively evaluate the effects of policies and interventions is to answer the prospective question of what effects they would have if they were to be endogenously adopted by a government, organization, or other actor. Yet these forms of agency are precisely what researchers seek to abstract away from when we substitute exogenous variation for endogenous choice in our research designs. In doing so, we make the implicit assumption that the effects of an action are the same regardless of whether that action has been exogenously triggered or endogenously chosen. How consequential is this assumption? Are there circumstances in which it might not be valid?

This review argues that some causal processes are inherently endogenous, in that they pertain to actions whose effects cannot be understood independently of the decision process that led them to be undertaken. In other words, a causal process is inherently endogenous when the effects of an actor's action (i.e., an independent variable) are conditioned by the fact of the actor having endogenously chosen to take that action, as opposed to having the action exogenously imposed upon them. In such cases, achieving an unbiased estimate of the action's direct effects requires cutting out the endogenous choice aspect of the causal process. But this changes the causal process being studied and thus the estimand. Studying inherently endogenous causal processes by leveraging observational or experimental variation can therefore give us an unbiased estimate of an action's causal effects, but only at the cost of forcing us to ask a slightly different question than the one we actually want to understand.

Faced with such inferential challenges, many researchers opt to frame their research questions in ways that abstract away from the inherently endogenous aspects of causal processes—even if these are central to the phenomena being studied and relevant to real-world decision-makers—or simply avoid studying them altogether. While many scholars do conduct empirical research that centers inherently endogenous processes and variables, they often do so outside of the framework of causal inference. Indeed, the issue of whether and how to study inherently endogenous causal processes is arguably at the center of a growing rift within political science and other disciplines that risks fragmenting scholars into two distinct epistemic communities: one that feels it can frame questions only in terms of causal inference and thus seeks out exogenous sources of variation and explanation, and another that rejects causal inference, considering it a straitjacket that inhibits the asking of important questions about choice, agency, and complex social phenomena. This methodological rift overlaps in part with a set of earlier fault lines and unresolved debates around the relationships between structure and agency, exogenous and endogenous social change, and determinism and indeterminism (cf. Riker 1984, Levi 2006). In naming and conceptualizing inherently endogenous causal processes, this article therefore aims to neither resolve these debates nor critique causal inference methods per se, but rather to provide a common conceptual framework and language to help bridge this methodological and epistemological divide and enable more constructive conversations among scholars.
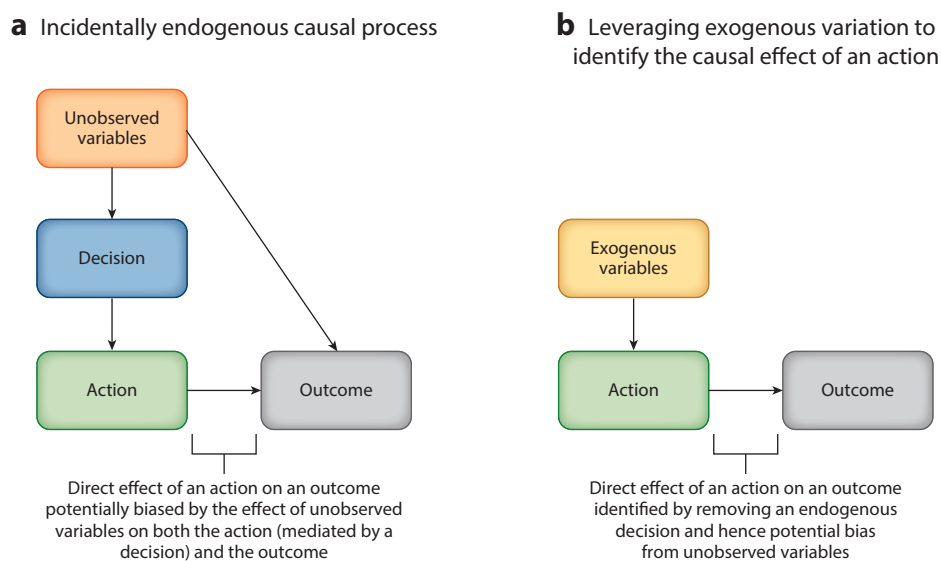
I begin by briefly laying out a framework that summarizes current thinking about exogenous and endogenous variables in causal processes, how quantitative and qualitative causal inference

methods can deal with problems of endogeneity, and what distinguishes endogeneity problems that are incidental (i.e., solvable with appropriate research design) from those that are inherent to the causal process itself (i.e., unsolvable). I then presents three causal mechanisms through which inherent endogeneity can arise: (*a*) when decision processes transform actors, (*b*) when decisions are based on private information that also moderates the effect of the subsequent actions, and (*c*) when decisions function as signals in strategic interactions. These are illustrated with empirical examples drawn from a range of literatures, such as institutional reform, community natural resource governance, and the determinants of interstate conflict.

The final part of the article addresses the question of how we can learn about inherently endogenous causal processes, if standard causal inference methods on their own are insufficient. I argue that this requires researchers to treat actors' endogenous decisions as integral parts of the causal processes we are interested in studying, rather than trying to find ways to abstract away from them. I survey a range of methods for doing so from both positivist and nonpositivist research traditions. These range from causal inference designs that do not rely on exogenous variation to formal game-theoretical modelling and noncounterfactual-based studies of mechanisms, interpretivist and constructivist approaches to understanding how actors make choices and how choice processes and mental models shape the structure of causal processes, and complex systems approaches that embrace indeterminism in causal processes. While these approaches have often been viewed as rival alternatives and do differ in the kinds of questions they ask and answers they give, combining them presents novel opportunities to arrive at richer understandings of inherently endogenous causal processes.

INCIDENTAL ENDOGENEITY PROBLEMS AND CAUSAL INFERENCE

Figure 1*a* depicts the classic setup of the types of endogeneity problems that causal inference methods are designed to overcome. In it, an actor makes a decision to take an action (i.e., an

a Incidentally endogenous causal process

b Leveraging exogenous variation to identify the causal effect of an action

Direct effect of an action on an outcome potentially biased by the effect of unobserved variables on both the action (mediated by a decision) and the outcome

Direct effect of an action on an outcome identified by removing an endogenous decision and hence potential bias from unobserved variables

Figure 1

Incidental endogeneity and causal inference. (*a*) Depicts how unobserved variables can bias the estimated effect of an action on an outcome. (*b*) Shows how causal inference methods use exogenous variation in the action to cleanly identify this effect.

intervention, policy, treatment, etc.) that may affect an outcome. Attempts to identify the direct effect of the action on the outcome are confounded because there may be unobserved variables that affect both the outcome and the endogenous decision to take the action. Within a statistical framework, this causal endogeneity would manifest as potential statistical endogeneity (i.e., a correlation between the independent variable and the error term). Many causal inference research designs therefore aim to identify the direct effect of an action on an outcome by examining cases where the determination of the action is driven not by an endogenous decision but by some other exogenous variable that can be observed (**Figure 1b**).

The most conceptually simple way to achieve this is by randomizing treatment, which allows the researcher to avoid statistical endogeneity by imposing causal exogeneity—by cutting out the endogenous upstream decision to take the action. This can also be achieved observationally through natural experiments or instrumental variables methods, in which the researcher identifies a causal factor that is exogenous to the actor being studied but that nevertheless affects the determination of the action variable and that does not have a direct impact on the outcome (other than through the action). If an appropriate exogenous variable can be found then researchers can estimate the direct effect of an action on an outcome, even in cases where the actor also makes endogenous decisions, by focusing their estimation only on the variation in the action that is due to variation in the exogenous variable. Similarly, in qualitative research, scholars often use most-similar case-comparative designs to focus on a variable that is exogenous to the actors being studied—for example, some historical event or factor—and trace how that exogenous variation leads to different decisions, actions, and outcomes in the different cases.

Causal inference methods that leverage exogenous variation thus treat endogeneity as an incidental problem that can be solved, at least if a researcher can find a case where suitable exogenous variation can be found or induced. If we can estimate the direct causal effect of the action on the outcome in one case or sample, then we can use that to generalize to other cases or samples where we cannot estimate causal effects because there happen to exist problems of incidental endogeneity, subject to the usual external validity concerns around the similarity of contexts and populations (cf. Williams 2020, Findley et al. 2021, Slough & Tyson 2024).

In this view, endogeneity is primarily a nuisance, a methodological problem that we can overcome if we are clever enough with our research design. The tacit assumption underlying this approach is that the endogenous decision depicted in **Figure 1a** is epiphenomenal, something that we can safely abstract away from without losing any understanding of the underlying causal relationships. More precisely, the assumption is that the direct effect of an action on an outcome conditional on an endogenous decision to take that action having been made is the same as the unconditional direct effect of the action on the outcome (or the direct effect conditional only on exogenous variation). In other words, it assumes that the decision process itself does not matter for the outcome of the action to which the decision leads. The next section explores causal mechanisms that might lead to violations of this assumption, in which the endogeneity is not incidental but inherent.

THREE MECHANISMS OF INHERENT ENDOGENEITY

This section lays out three theoretical mechanisms through which inherent endogeneity could arise in causal processes and discusses their implications for the validity and interpretation of estimates from common causal inference methods. In doing so, I focus on mechanisms of endogeneity that are upstream of the actions being studied, in the decision processes that lead to them. This is distinct from the growing stream of research within the causal inference tradition on general equilibrium and long-term effects (cf. Acemoglu 2010, Stevenson 2023), which focuses

on modelling and understanding downstream endogeneity in other actors' reactions to an initial action but nevertheless typically still relies on exogenous variation in the initial action.

The common theme across these three mechanisms of inherent endogeneity is that they represent causal processes in which the reason why an actor took a given action matters for the outcome of that action. Put another way, a causal process is inherently endogenous when the decision to undertake an action is an inseparable part of the mechanism(s) through which the action leads to an outcome.

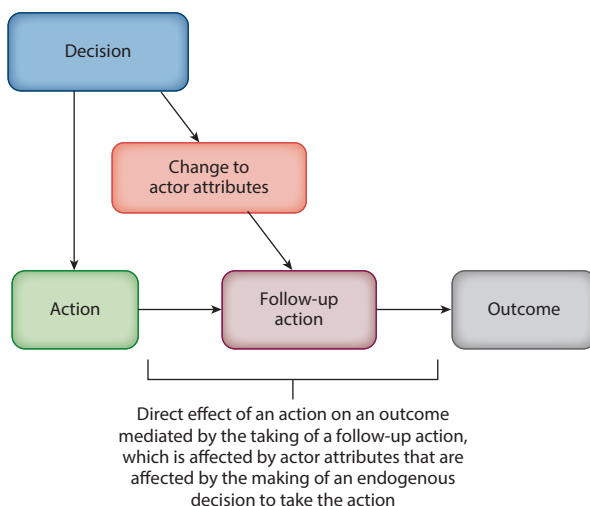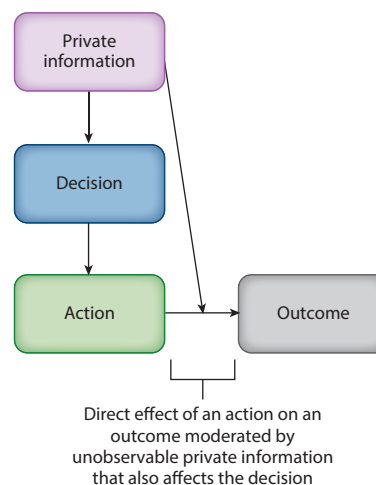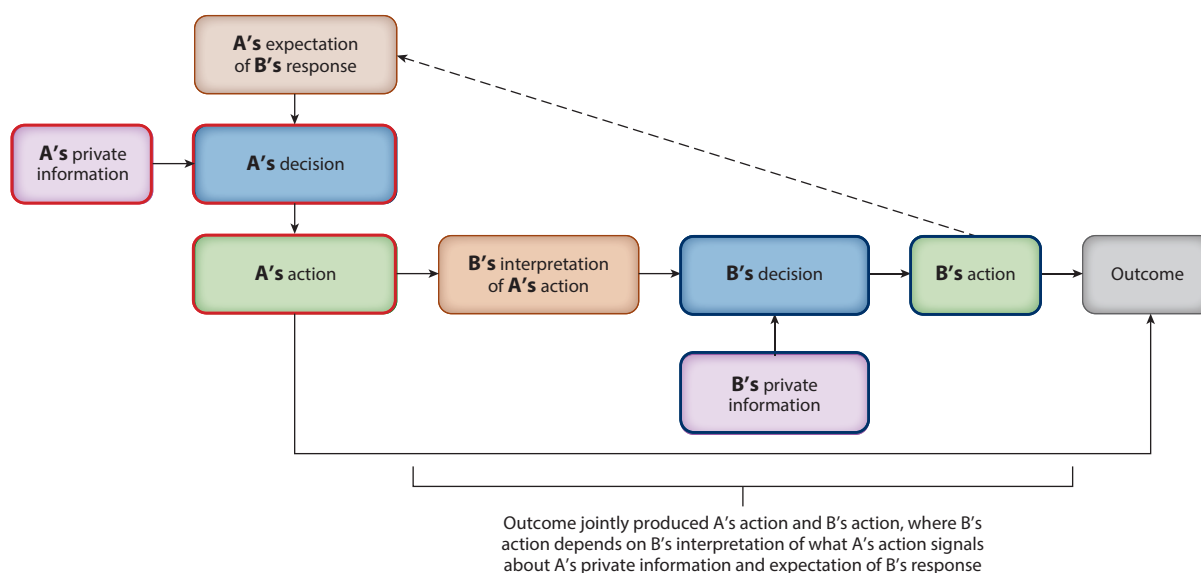## Mechanism 1: Decisions Transform Actors

Endogenous decisions may matter not only because they lead to a certain action but also because the process of making a decision could itself change the actor(s) involved in ways that might affect the taking of follow-up actions that mediate the outcome of the action (**Figure 2a**). These changes could affect individual characteristics, such as information, preferences, cognitive frames, or mindsets, as well as group characteristics, such as relationships, common understandings, or trust. If the phenomenon researchers wish to learn about has the causal structure depicted in **Figure 2a**, then leveraging exogenous variation in the action (as per **Figure 1b**) would miss the part of its effect that is driven by the *Decision → Change to Actor Attributes → Follow-Up Action* pathway. The researcher could still observe whether actors took downstream follow-up actions in response to the externally imposed action and thus achieve internal validity, but the missing causal pathway means that the researcher would not know whether these follow-up actions would have been different if the action had been endogenously chosen.

This mechanism could potentially exist across a wide range of political and social phenomena. Consider a government bureaucracy that is considering whether or not to undertake a reform—adopting a new set of performance targets and associated accountability measures, say. The process of deciding whether to adopt the reform and how to design it is a matter not only of creating the technical design of the reform but also of building trust and relational understandings among managers and other key stakeholders, of creating a sense of shared ownership, and of laying out and internalizing the follow-up steps and expectations of future behavior that will be necessary for the successful implementation of the reform.[1]

The literature on the reform of public institutions presents many examples that are consistent with the idea that endogenous decision processes may both directly affect final outcomes and moderate the impacts of the reforms that were adopted. In their study of the impact of countries' participation in the Open Government Partnership, Piotrowski et al. (2022, p. 23) distinguish between the "direct" effects that arise from the policy commitments that participation entails, which were "limited, context dependent, and superficial," and the "indirect" effects of participation in terms of the formation of new national and transnational networks, normative changes in the actors themselves, and the building of reform coalitions, which were more consequential in the long term. Williams (2026) studies large-scale civil service reforms and observes that they are inherently endogenous in that they involve the taking of endogenous decisions that plausibly affect the subsequent implementation of reforms, either positively or negatively. Mangla's (2022) study on "deliberative bureaucracy" shows that successful policy implementation

---

[1]In a related discussion, Paul & Healy (2018) explore the inferential difficulties associated with evaluating "transformative treatments" that not only affect a person's life but also might change their preferences, such as becoming a vampire or having a child. The authors focus mainly on how this complicates ex post efforts to measure subjective feelings about the action, even if randomly assigned, rather than on the idea that a transformation of actors resulting from endogenous choice might lead to different follow-up actions and thus to different outcomes.

a Endogenous decisions transform actors

Direct effect of an action on an outcome
mediated by the taking of a follow-up action,
which is affected by actor attributes that are
affected by the making of an endogenous
decision to take the action

b Private information moderates the effect of an action

Direct effect of an action on an
outcome moderated by
unobservable private information
that also affects the decision

c Endogenous decisions as signals in strategic interactions

Outcome jointly produced A's action and B's action, where B's
action depends on B's interpretation of what A's action signals
about A's private information and expectation of B's response

Figure 2

Three mechanisms of inherent endogeneity. (*a*) Illustrates how endogenous decisions can transform actors in ways that affect other downstream actions. (*b*) Illustrates how actors' private information can moderate the effects of their actions. (*c*) Illustrates how endogenous decisions can act as signals and thus influence other actors' responses to an action.

in India's education sector came not from strict compliance with top-down rules and dictates but rather from iterative interpersonal coordination and communication that created shared understandings, both within the bureaucracy and between the bureaucracy and actors in society. The aid effectiveness literature has emphasized the importance of local ownership of reforms for

enabling consultation and coalition-building among stakeholders (Tsikata 2001) and because of its potential psychological effects (Repucci 2014). In the private sector management literature, Gibbons & Kaplan (2015, p. 449) quote an executive whose firm had implemented the popular balanced scorecard approach to measuring organizational performance:

> You could take our scorecard and give it to a competitor and it wouldn't work. You had to have sweated through the hours and hours of work and effort that went behind the card to get the benefits from the measures. That's what brings it to life. It's got to become part of the company's belief system, almost a religion—the benefits don't come just from having a piece of paper with a scorecard on it.

In a similar vein, Mansuri & Rao (2013) distinguish between "organic" and "induced" forms of participatory governance, noting that the latter tend to be more transformative. In their method-ological review on comparative-historical scholarship, Thelen & Mahoney (2015, pp. 26–27) note that "what also matters is the process through which these institutions themselves emerge. The endogenous political processes through which democratic institutions arrive…as much as the institutions themselves, are what produce the ultimate effects—through the way these processes transform citizen expectations and reconfigure social and political dynamics." In the public health literature, Abimbola (2019) contrasts the "surgical" perspective (in which health system change happens due to external intervention) with the "organic" perspective (in which change happens from within). He observes that "it is much easier to see 'surgical' change…than it is to see 'organic' change" (Abimbola 2019) and that this bias in research methodology has been amplified by the historical predominance of "the foreign gaze" over "the local gaze" in what research gets funded and published and where perceived agency gets located.

In lab experimental work in economics, Dal Bó et al. (2010, p. 2206) demonstrate that subjects playing a collective action game behave more cooperatively when operating under endogenously chosen rules than when operating under externally imposed ones, even controlling for both the content of the rules and the informational signals conveyed by actors' choices of preferred rules, "so that a democratically selected policy will not have the same effect when imposed undemocratically." They reflect on the implications of their findings:

> Much applied work in economics seeks to identify the treatment effect of policies, institutions, or products. Since people usually choose their policies, institutions, and products, it is necessary to ac-count for selection into treatment to measure the "true" treatment effect (i.e., one that does not reflect selection). Based on such estimates, policy recommendations may be made to assign the treatment with-out choice (that is, exogenously). Our experimental results suggest that such policy recommendations may be unwarranted given that the treatment effect may differ based on whether it is exogenously or endogenously determined, even after controlling for selection. (Dal Bó et al. 2010, p. 2206)

Outside of the lab, the literature on community natural resource governance schemes is an-other example of these conceptual and methodological challenges. The huge academic and policy interest in such schemes arose precisely because they represent self-organized, decentralized mechanisms for solving common pool resource problems (cf. Ostrom 1990, Agrawal et al. 2023), but they also present a classic endogeneity problem: Some of the same factors that lead some com-munities but not others to create and sustain resource governance mechanisms might also affect the outcomes of interest, so we cannot disentangle selection and treatment effects. The Metaketa III initiative sought to shed light on these issues by conducting six harmonized randomized controlled trials (RCTs) in different countries, in which nongovernmental organizations provided data and technology to lower the cost of monitoring and thus encourage community-driven resource governance, with broadly positive effects (Slough et al. 2021). Turning community monitoring into an intervention that could be studied by field experiments thus allowed the researchers to cut out the upstream endogeneity (i.e., the selection problem) associated with

self-organized institutions, leaving only downstream endogeneity in take-up of the external provision of information (which could be measured by comparison against a counterfactual). But it also changed the causal process under study, from self-organized monitoring to "monitoring supported by external actors" (Slough et al. 2021), cutting out—or at least altering—the second potential pathway of impact in **Figure 2a**. Of course, the field experiments are still informative about the first pathway of impact, and the authors themselves are admirably transparent and thoughtful about these distinctions (and also note that the effectiveness of externally induced community natural resource monitoring is itself a question of significant policy interest). Still, even this close-to-best-case set of well-executed, harmonized, multisite field experiments illustrates the challenges that researchers and readers face in trying to use causal inference research designs to learn about real-world phenomena that might be characterized by inherent endogeneity.

Mechanism 2: Private Information Moderates the Effect of an Action

A second mechanism through which inherent endogeneity might arise is when decision-makers' unobservable private information moderates the effects of their actions, as in **Figure 2b**. In such situations, replacing an endogenous decision with an exogenously driven one might fail to capture the heterogeneity in the action's effects that is driven by the private information that guides the endogenous decision. Exogenously imposing an action (as per **Figure 1b**) might allow a researcher to recover an unbiased estimate of the direct effect of the action, but often the quantity of real-world interest (particularly for prospective decision-making) is the effect of the action conditional on the private information of the decision-making actor(s).

To see how such a mechanism could arise, let us return to the empirical examples from the previous section. A policymaker considering whether to undertake an institutional reform may be influenced in their decision by the level of political support they believe they have (the private information), and that political support may also matter for the enforcement, implementation, or public acceptance of the policy reform (the moderating effect). Similarly, in collectively choosing whether or not to establish a community natural resource government scheme and (if so) how to design it, each community member's decision might be influenced by their own private knowledge about their own type or goals as well as their perceptions of the level of trust within the group, and these factors may also moderate the effectiveness of the scheme.

In both situations, an appropriate research design relying on exogenous variation in the action could recover an unbiased estimate of the direct effect of the action, but not the moderating effect of the actors' private information. If one defines the estimand of interest simply as an unbiased estimate of the direct effect, then this is not a problem. But if the researcher is interested in understanding theoretical mechanisms, or in providing evidence that future decision-makers can use to prospectively predict the effects of undertaking the same action, then omitting the moderation causal pathway might be consequential.

How plausible is it that this causal pathway might be important for some phenomena? Allcott (2015) studies an intervention intended to reduce household energy use (mailing a personalized energy report) across 111 RCTs across different sites in the United States and finds that impacts in the first 10 sites were significantly higher than impacts in the remaining sites because "early sites were strongly positively selected from later sites through mechanisms associated with the treatment effect" (pp. 1119–20). Allcott also documents similar evidence of site selection bias—both positive and negative—in microcredit RCTs and clinical trials. Of course, the idea that interventions are heterogeneous across contexts is nothing new, and indeed some of the factors driving heterogeneity and positive selection in Allcott's context were observable. But simply controlling

for these observables and/or conditioning treatment effects on them would not have resolved this issue since a large share of the heterogeneity in effectiveness both within and across sites was accounted for by unobservables (which were uncorrelated with the observables that drove selection), despite the unusually rich data that Allcott's context afforded.

An analogous issue could arise with respect to nonexperimental causal inference scenarios, such as instrumental variables methods. The treatment effects estimated are local average treatment effects that pertain to the subset of "units that receive treatment if and only if they are induced by an exogenous IV [instrumental variable]" (Aronow & Carnegie 2013), which may well be different than the subset of units that endogenously opt into treatment based on their private information. So, **Figure 2b** could be thought of as capturing a special case of the more general and much-debated challenges in recovering average treatment effects (ATEs) from local average treatment effects (LATEs) (cf. Deaton 2010, Heckman & Urzua 2010, Imbens 2010); while a range of econometric methods exist to do so under certain conditions and assumptions (as discussed later in this review), this is far from universally possible. If these variables could be perfectly measured or proxied, then they could be dealt with econometrically using standard methods for estimating heterogeneous effects. But if they are not just unobserved (i.e., happen to not be measured) but also are fundamentally unobservable because they relate to pure private information such as interior mental states, perceptions, or strategies, then at best they can be imperfectly proxied.

This mechanism of inherent endogeneity thus recalls the previous section's observation that why an actor chooses to undertake an action might be important for that action's effects. This could be because that endogenous decision triggers downstream follow-up actions, as the previous section discusses, but this section points out that it could also be because the rationale behind the endogenous decision is driven by upstream private information. The next section explores these issues further, considering multi-actor phenomena for which actors' real and perceived goals, strategies, and information are (*a*) consequential for understanding the effects of their actions and (*b*) hard to proxy, not least because actors often have reasons to not divulge or even to mislead about them.

Mechanism 3: Endogenous Decisions as Signals in Strategic Interactions

Figure 2c depicts a third mechanism of inherent endogeneity, this time with two actors, A and B, who each make decisions over actions that jointly produce an outcome that yields payoffs to both A and B. A's decision of what action to take depends on two sets of factors: private information that A holds (e.g., about A's type, the state of the world, or A's strategy) and A's expectations about how B will react to A's actions. After A makes a decision and acts, B observes this action and interprets what it thinks the action signals about A's private information. B then decides on its own action based on this interpretation combined with its own private information, and A's and B's actions together determine the outcome. (If the game were a repeated one, then B's action in the first period would feed back into A's expectations in the following period, and similarly for B's interpretations.) This basic mechanism of strategic interaction nests many types of sequential games and can also be generalized to simultaneous repeated games.

Within this causal structure, attempts to estimate an average treatment effect of A's action face an obvious problem of endogeneity, so a researcher might try to identify or induce exogenous variation in A's action by finding an instrumental variable that drives A's action or by getting A to randomize its decision. A researcher could then estimate an unbiased average treatment effect of A's action on B's action and thus on the outcome. But if A's decision is driven by exogenous variation rather than an endogenous choice, then it no longer serves as a signal of its private information.

B's interpretation of A's exogenously driven action would thus differ from B's interpretation of A's endogenously chosen action, in turn potentially leading B to choose a different action. So, the treatment effect estimated by the researcher when A's action was exogenously determined would not necessarily apply to the case where A chooses its action endogenously, because B would react differently if it knew A was making an endogenous decision rather than having an action exogenously imposed on it.

Where might we encounter phenomena with this causal structure? The bargaining theory of war presents one example. This approach hypothesizes that conflict emerges from strategic interactions as a result of uncertainty over the private information held by each party (e.g., military capabilities, strategies, level of political resolve), which inhibits parties from making and sustaining mutually beneficial compromises (cf. Morrow 1989, Fearon 1995). Yet despite its widespread theoretical acceptance, it has proven far more difficult to provide direct empirical support for the model, with most empirical work being conducted either in lab or survey experiments or through studying indirect implications of the theory (Gartzke & Poast 2017). From a causal inference perspective, the challenge is to understand the consequences of actions taken by A (e.g., alliance forming, public rhetoric, deterrence actions) on reactions by B (e.g., escalation or de-escalation) and hence on the mutually produced outcome (e.g., war or peace). This is obviously difficult for practical reasons, but suppose that exogenous variation that drove A's actions but was uncorrelated with A's private information could somehow be found or induced. This exogenous variation would indeed allow researchers to estimate an average effect of A's actions that was unbiased by A's unobservable private information. But the fact that A's actions in these instances were being driven by exogenous variation would mean that the action would not serve as a signal of A's private information to B, so B would interpret and respond differently than if B knew that A's action was a purely endogenous choice based on private information rather than on the exogenous variable or experiment. Therefore, any research design that would produce an unbiased estimate would do so only for a different causal process than the one we are actually interested in understanding. As Gartzke (1999, p. 582) notes in his article titled "War Is in the Error Term," the uncertainty and indeterminism (and hence actors' agency in interpretation and action) in the bargaining theory of war is not a statistical problem incidental to efforts to empirically test the theory but is actually "the key causal variable" in it.

Bueno de Mesquita & Tyson (2020) capture another version of this mechanism in their discussion of the empirical challenges in estimating the causal effects of protests and governments' responses to them. They observe that many treatment behaviors (i.e., A's action) affect outcome behaviors (i.e., B's action) through two channels: the direct effect of the action and the informational signal about private information conveyed by A's choice of action. Causal inference research designs might approach estimating the unconfounded direct effect of protest by using some exogenous shock, such as the weather, to instrument for the size of protests. But Bueno de Mesquita & Tyson (2020, pp. 376–77) point out that "when governments observe a large protest on a sunny day, they know that it conveys different information than a large protest on a rainy day....Perhaps, the outcome agent, having observed the shock to treatment behavior, doesn't update her beliefs at all ('I know those extra 10,000 people are here just for the sun')." Thus, "the effect of changes in protest behavior estimated from weather shocks is an inherently different quantity than the effect of changes in protest behavior due to, say, unobservable changes in citizens' antigovernment sentiment or even unobservable shocks to the cost of protesting" (Bueno de Mesquita & Tyson 2020, p. 376). The authors construct a formal proof to show that even an ideal experiment does not capture the real-world estimand of theoretical interest, which is the combined direct and informational effects of an endogenously chosen action. They also show how the same issue applies to the literatures on the effects of electoral victories by female

candidates and the effects of repression of rebel groups by counterinsurgents. More generally, they argue that this "commensurability problem" is pervasive to any effort to estimate "the effect of behavior on behavior" when the choice of behavior serves a signaling function (Bueno de Mesquita & Tyson 2020, p. 375). The only two empirical examples the authors discuss in which it is possible to identify the direct effect of the action without bias from an informational channel both pertain to the rare circumstance in which A is acting randomly (or at least with an element of exogenous variation), but B does not know this and thinks A is acting purely endogenously. This further illustrates the broader point that why actors choose certain actions (and sometimes also other actors' perceptions of these reasons) matters for the outcomes of these actions.

Of course, understanding "the effect of behavior on behavior" describes many (if not most) of the phenomena in which political scientists—as well as sociologists, economists, psychologists, and other social scientists—are interested. In the case of institutional reforms discussed above, for example, how a public service union decides to respond to a management reform might depend on their perceptions of the government's underlying goals or strategies as well as of their own strength, which could be one reason (among many) why large-scale government policy initiatives frequently have weaker effects than small-scale pilot interventions led by researchers or nongovernmental organizations. What makes this even more challenging to study empirically is that the actors involved often have reason to mislead other actors (as well as researchers) about their private information. This mechanism may thus operate across a wide range of strategic interactions, and it seems likely to be particularly problematic when the actors themselves are conscious of the signaling function of their actions.

The three mechanisms of inherent endogeneity discussed in this section are conceptually distinct but share one crucial feature in common: The exercise of agency in decision-making processes affects the outcomes of the resulting actions. Inherently endogenous causal processes are thus somewhat analogous to bundled treatments, consisting of both an action and a decision that moderates the outcome of that action. Causal inference research designs that leverage exogenous variation work precisely because they eliminate this element of agency—that is, they unbundle the decision and the action—and thus achieve unbiasedness, but at the cost of subtly changing the causal process under study. How consequential this shift is will differ across different empirical phenomena and contexts, but there are ample theoretical and empirical grounds to suggest that scholars need to take it seriously.

It is tempting to think of inherent endogeneity issues through the lens of external validity, especially as political scientists have increasingly included construct validity as one criterion that might affect the comparability and hence generalizability of causal inference estimates across contexts (cf. Findley et al. 2021, Slough & Tyson 2023). Instead, I suggest that researchers are better off thinking of inherently endogenous causal processes as related but distinct to their exogenous or incidentally endogenous cousins: Yes, they share the same action, but the mechanisms through which the action affects the outcome are fundamentally different. So, what we want is not merely to try to generalize or extrapolate causal inference estimates to endogenously chosen instances of that action, but rather to understand endogenous actions on their own terms.

To do so, we must focus our attention on the upstream decision-making processes within which endogenously chosen actions, treatments, or independent variables are embedded and on the causal pathways that link them to downstream actions and outcomes. Understanding these connections is the key to learning about inherently endogenous causal processes and to reconciling rigorous causal inference with the theoretical recognition that human agency and endogenous choice often matter for the phenomena in which we are interested.

LEARNING ABOUT INHERENTLY ENDOGENOUS CAUSAL PROCESSES

The central challenge of studying inherently endogenous causal processes is that decisions and decision processes may affect the outcomes of the actions they lead to. Rather than seeking to abstract away from endogenous decision processes, scholars therefore need to find ways to better understand the combined, interactive, sometimes iterative effects of decisions and actions.

This section discusses a range of methods researchers might use to do this. These brief discussions are inevitably incomplete and selective. Instead, this section aims to highlight both the complementarities and tensions among these approaches and to point interested readers toward more comprehensive reviews, with the aim of facilitating greater dialogue and collaboration across different methodological and epistemological traditions.

First, thinking of endogenous decisions as part of the causal process through which actions lead to outcomes (rather than just as incidental sources of endogeneity) might lead us to think differently about two classes of causal inference estimators—difference-in-difference and regression discontinuity designs—that allow researchers to study actions that have been endogenously chosen. Difference-in-difference designs can be used to compare outcomes in units that have opted into a treatment with outcomes in units that have not, and so they seem to be good candidates for studying the effects of actions for which endogenous decision-making processes might matter. However, similar to the heterogenous treatment effects issues discussed in the previous section with respect to instrumental variables estimates, difference-in-difference estimators typically return average treatment effects on the treated (ATTs) rather than average treatment effects (ATEs) since units may have selected into treatment based on factors that also moderate the effect of the treatment. (Indeed, this is definitionally true in the mechanisms of inherent endogeneity depicted in **Figure 2b,c**). Generalizing estimated treatment effects to nontreated units is therefore problematic. At the same time, the estimand in difference-in-difference designs with endogenous selection into treatment is usually not the direct unconditional effect of the action on the outcome but is instead the effect of an action conditional on an endogenous decision to undertake the action, which is precisely the estimand of interest for inherently endogenous processes defined earlier in this article. In such instances, what we are studying is actually the combination of an endogenous decision and subsequent action, rather than the action alone; endogenous decisions could conceivably both moderate the effect of the action and have their own direct causal effect on the outcome.[2]

Regression discontinuity designs, on the other hand, can disentangle the effect of the decision process from the action by allowing agents to undergo endogenous selection-into-treatment decision processes but then inserting as-if-random variation between the decision and action stages using a cutoff.[3] For example, Clemens & Tiongson (2017) use a language test score cutoff to study the effects of migration on Filipino workers who applied to work temporarily in South Korea. This allows the researchers to net out any direct or moderating effects of endogenously

---

[2]This raises the interesting implication that the effect of an action estimated by a difference-in-difference estimator might be different than the effect of the same action measured instead through an RCT or instrumental variable, if one of the mechanisms of inherent endogeneity applies to the underlying causal process.

[3]Difference-in-difference design can sometimes also capture this by using as counterfactuals units that decided to take the action but were somehow prevented from doing so, separating the effect of the decision from that of the action (although whatever factor intervened to prevent the action might also have its own effects, so the comparison might not still be clean). The more general point is that causal mechanisms can differ between treatment and control groups in complex and interesting ways, and researchers need to think carefully about which of these mechanisms such designs are and are not capturing and netting out in different situations.

opting into treatment, thus yielding the direct effect of the action alone. But these estimates of the effect of treatment cannot necessarily be generalized to agents that did not opt into treatment since the estimates are of direct effects conditional on having decided to try to obtain treatment rather than direct unconditional effects. There may thus be scope for researchers to use these differences in whether and how endogenous decisions are captured by different causal inference methods to distinguish between and disaggregate the effects of these different causal pathways.[4]

Indeed, some studies have already done something akin to this in experimental settings by manipulating the nature of upstream decision processes offered to agents and using the difference in outcomes across endogenously chosen versus exogenously imposed actions to measure or control for the effects of endogenous choice. This has been done in lab settings (e.g., Dal Bó et al. 2010, Cason et al. 2025) and in field experiments (e.g., Leaver et al. 2021). But while re-randomization and other forms of experimentally varying agents' choice processes are one potential way to obtain cleanly identified estimates of the impacts of endogenous choices, practical and sometimes ethical challenges limit the potential scope for this technique to be used in studying inherently endogenous processes in many real-world contexts.

Another approach to the recognition that endogenous decisions might matter is to use formal modelling of upstream decision processes to better understand the interpretation of counterfactual-based inference about the downstream effects of the resulting actions. Bueno de Mesquita & Tyson's (2020) work on decomposing the direct and informational effects of popular protest on government response is perhaps the best example of this. While their main conclusion is that the circumstances under which these two channels actually can be distinguished are very rare, their analysis nevertheless contributes to a significant reinterpretation of a large empirical literature. Game-theoretical formal modelling of upstream decision processes and their effects has also long been central to other fields of inquiry, such as the analytical narratives tradition in political science (Bates et al. 1998) and the study of organizational decision-making and relational contracts (cf. Chassang 2010, Gibbons et al. 2013), although these have relied mainly on noncounterfactual-based qualitative and quantitative evidence in their empirical applications.

Another recent stream of work within the causal inference tradition has been on the scaling and implementation of interventions (cf. Al-Ubaydli et al. 2017, Bold et al. 2018). The main motivation behind this literature—a desire to understand the frequently observed disparity between the effects of interventions in controlled conditions versus real-world, large-scale conditions—certainly connects with the motivation behind the study of inherent endogeneity. However, studies in this vein have mainly tackled these challenges by experimentally varying properties of the implementing agent or intervention, rather than by investigating the potential role of endogenous decision processes or the iterative sequencing of decisions and actions in implementing and sustaining interventions. In contrast, the extensive literatures on the diffusion of public policies and management practices (cf. Tolbert & Zucker 1983, Shipan & Volden 2008, Ansari et al. 2010) place endogenous decision processes front and center, but usually without empirically exploring how decision motivations or processes might drive differential outcomes. The literature on site selection bias (Allcott 2015) and designing research in order to minimize it (Gechter et al. 2023) is perhaps the best example of an empirical literature that combines inquiry into the adoption of interventions with an assessment of how adoption affects outcomes, although the authors also note that there are relatively few interventions for which there exist a large enough body of studies to be able to study it empirically.

---

[4]Relatedly, Kowalski (2023) reviews methods for researchers to use patterns of (non)compliance and take-up within experiments to study external validity, although these methods mainly pertain to downstream responses to policies rather than the type of upstream decisions that are the focus of this article.

Qualitative methods for deductively inferring causality, such as comparative-historical analysis (Mahoney & Thelen 2015), qualitative comparative analysis (Ragin 2014), process tracing (George & Bennett 2005), and other mixed-methods approaches (cf. Lieberman 2005, Glynn & Ichino 2015, Humphreys & Jacobs 2023), may also help researchers understand decision processes and their effects. First, these methods' greater analytical flexibility might allow researchers to observe and analyze decision processes—which tend to be hard to quantify—in more nuanced ways. Second, richer information on decision processes can help researchers select more comparable cases for comparison—perhaps even critical junctures (Capoccia 2015)—where similar endogenous choice processes and structural conditions led to different actions whose outcomes can then be compared. Third, and perhaps most importantly, these methods often center issues of temporality (Thelen & Mahoney 2015) and conceptualize causal processes as sequences of iterative decisions and actions rather than one-off interventions or variables, which helps to identify and study the mechanisms of inherent endogeneity captured in **Figure 2a,c**. At the same time, such nonexperimental comparisons will always be subject to incidental endogeneity concerns around selection on unobservables, and within-case inference will always be contestable except in the rare cases of obvious smoking gun causal links. The management literature on longitudinal process studies in organizations (cf. Pettigrew 1990, Tsoukas & Chia 2002) offers an example from outside of political science of both the potential and challenges of adopting a temporal, process-based analytical perspective: It opens rich opportunities for nuanced empirical analysis and theory-building but also tends to be short on conclusive and replicable empirical findings. So, while these varied methods can certainly help us learn about inherently endogenous causal processes, they cannot resolve the underlying inferential challenges discussed above.

Another approach to studying inherently endogenous causal processes is to give up on trying to make strong causal claims and focus instead on rich description of both upstream decision processes and the downstream mechanisms associated with them. For example, Lowande & Salis-Muñiz (2025) propose a new theory of why political leaders limit bureaucratic discretion based on their analysis of US presidents' decisions about adopting abortion bans in foreign aid. Key to their analysis is their use of archival documents and interviews with decision-makers, which allows them to observe decision-makers' strategic considerations and "to study prohibitions that were proposed but not adopted" (Lowande & Salis-Muñiz 2025, p. 649). Williams (2026) studies more than 100 system-wide bureaucratic reforms across six countries in Africa for which endogenous adoption makes strong causal inference impossible, focusing not on identifying treatment effects but on abductively building theory about "mechanisms of success" and "mechanisms of failure" based on documenting repeated empirical patterns and examining their consistency with different potential underlying causal models. A key mechanism that emerges is that whether downstream reform actors take implementation actions depends on the credibility of the reform, which in turn stems partly from the nature of the endogenous decision process and motivations behind it. In addition, Piotrowski et al. (2022) explicitly trace the indirect effect mechanism of the global transparency initiative they study back to countries' endogenous choice to enter it.

It is notable that the three examples in the preceding paragraph each rely heavily on expert interviews. von Soest (2023, p. 277) observes that "experts are the observers of and mechanics behind what social scientists call 'causal mechanisms,'" so it makes sense that they would be especially valuable for understanding the decision–action–outcome linkages that characterize inherently endogenous processes. Of course, other types of qualitative and quantitative data and methods can also help scholars better understand mechanisms, as discussed in a rapidly growing literature in political science and adjacent fields that is too vast to adequately review here (cf. Hedström & Ylikoski 2010, Mahoney 2010, Marchal et al. 2012, Imai et al. 2015, Goertz 2017, Bueno de Mesquita et al. 2021, Widner et al. 2022). A key debate in this literature is the extent to which

mechanisms can or should only be studied once a main effect has been causally identified, as opposed to noncounterfactual-based, data-consistent-with-theory styles to building and interpreting evidence. When we want to study inherently endogenous causal processes, however, we are often effectively forced into the latter camp—with all its limitations—since exogenous variation in the treatment variable of interest means that the mechanisms of inherent endogeneity do not exist, and vice versa.

So far, the approaches surveyed have all fallen squarely within a positivist research tradition, in the sense that they assume that the world is comprised of stable causal relationships among variables and that the goal of social scientists is to uncover these relationships, even if there are different perspectives about how to achieve that. But a range of other approaches to studying inherently endogenous processes depart from this paradigm in their epistemology, what they define as their research goals and questions, and thus their empirical methodologies.

Interpretivist approaches to political science research study decisions by investigating how actors make and interpret meaning in exercising their agency. Interpretivism is not a coherent body of theory or methods, but rather a broad family of approaches that center the study of actors' own understandings of what they are doing and why. It thus encompasses a very wide range of analytical strategies, from semistructured interviews to textual analysis to participatory action research (Yanow & Schwartz-Shea 2014), as well as methods such as political ethnography (Schatz 2009) that can help researchers better understand what drives policymakers' upstream decisions and how these decision processes are constitutive of further action by themselves or other actors. Interpretivism has an especially strong tradition in organizational sociology and behavior, with concepts such as sensemaking having long been central to understanding how individuals themselves perceive the causal structures within which they are embedded, and hence the causes and consequences of their actions (cf. Weick 1995, Sandberg & Tsoukas 2015). In studying education reform and delivery, for example, scholars have taken interpretivist approaches to understanding the role of agency in policy implementation, highlighting "individuals' power to make strategic choices and to disrupt, mediate, combine, and interpret in the implementation process" (Coburn 2016, p. 470). Similarly, Donaldson et al. (2021) showed how superintendents in high- versus low-performing districts interpreted their roles in implementing a statewide principal evaluation policy differently, leading them to focus on different aspects of the policy in their interactions with principals. While much of this research has focused on actions and decisions that are downstream of the high-level policy decisions whose implementation they study, they could be productively applied to the types of upstream decisions that are key to the mechanisms of inherent endogeneity identified in this article.

Whereas interpretivism tends to focus on trying to understand what is happening inside individual actors' minds, constructivism instead focuses on studying how decisions are affected by individuals' embeddedness in social, institutional, and intellectual contexts. Constructivism thus aims to understand how social norms, mental models, and logics of appropriateness develop, spread, and influence individual action. Within political science (and sociology), constructivist research has been especially prominent in international relations, where it has become established as a widely accepted and debated school of thought centered around studying how actors come to conceptualize their interests, internalize appropriate ways to behave, and construe causal connections between actions and outcomes (cf. Kentikelenis & Babb 2019, Finnemore & Wendt 2024). Other IR scholars have adopted a practice theory perspective on these questions, focusing attention on the role not just of abstract ideas but also of day-to-day practices and relations (Adler & Pouliot 2011). Outside of IR, scholars have taken constructivist perspectives to studying how economic reasoning came to be dominant in public policy discourse and practice over the second half of the twentieth century (Popp Berman 2022, Killick 2023). Networks in particular hold

significant potential explanatory power for exploring how individuals' identities, actions, and relationships are iteratively constitutive of each other (Padgett & Powell 2012). Taking a constructivist perspective also highlights the particular importance of universities (e.g., Abbott 2005) and academic journals (e.g., Abimbola 2018) due to their roles in knowledge production, training, and socializing elite actors. This literature helps highlight that while academics from causal inference traditions are used to thinking of themselves as sitting outside of the causal processes they study, in many cases research and researchers themselves have causal effects on the systems they aim to understand.

Finally, one might approach research design not as an effort to banish confounding in the pursuit of identifying stable, generalizable causal relationships between actions and outcomes, but rather as an effort to understand whether, when, and how the same action can often lead to different outcomes due to the intrinsic complexity of social and political phenomena. For example, in his review of Wawro & Katznelson's (2022) book on variation in causal relationships across time periods in historical analysis, Carpenter (2024, p. 606) emphasizes the limits that this places on comparison and generalizability: "If historical spaces, institutions, and processes are considered like equilibria (or like combinations of equilibria), then a basic problem of incommensurability arises. There are, put differently, entire histories and cultures that scholars often regard as not fundamentally comparable with one another in the sense of constituting a 'case' of a more general phenomenon." Another set of approaches, rooted in complex systems theory and polycentric governance, arrive at the similar conclusion that the causal effects of some actions are often effectively indeterminate or highly context-specific (cf. Thiel et al. 2019, Mansoor & Williams 2024). This critique—and the development of theory and empirical research to grapple with its implications—is perhaps most advanced in the health field (cf. Gilson 2012, Hawe 2015). Writing from within economics, Stevenson's (2023) review of criminal justice interventions also locates their persistent failure to meaningfully shift outcomes in the idea that changing a single variable in a complex system with self-reinforcing properties is often an ineffective way to drive meaningful social change (although her review focuses more on downstream endogeneity and general equilibrium issues than on upstream inherent endogeneity). To the extent that individual agency itself is impossible to fully predict based on measurable structural factors, the results of decision processes and their linked actions and outcomes may always remain impossible to perfectly predict or fully explain.

Taken together, the literature discussed in this section comprises a vast range of epistemological, topical, theoretical, and methodological approaches. However, they all represent ways that scholars can link the study of actors' decisions about what actions to undertake with the study of the effects of these actions. They thus can each shed different kinds of light on different parts of the mechanisms of inherent endogeneity described in the previous section. Overall, research would benefit from greater efforts to combine these ways of learning, and in general there are more complementarities and fewer fundamental incompatibilities than would be suggested from the limited existing engagement across the boundaries of many of these research traditions. But the combining of these approaches does also come at some cost to parsimony, both in terms of the types of research questions we ask and the types of answers we find. Augmenting a meta-analysis of the impacts of a particular education reform with a constructivist analysis of the varied reasons why actors adopt that reform, for example, would undoubtedly lead to a richer understanding of the reform's real-world impacts, but this richness would probably not manifest itself in tighter confidence intervals. On the other hand, once we recognize the need to combine them and devote more effort to trying to do so, perhaps we will realize that these approaches are even more commensurable than we currently realize and find new ways to make them so.

CONCLUSION

The two main themes of this review are that (*a*) inherent endogeneity poses significant challenges for the application and interpretation of causal inference methods in phenomena where one or more such mechanisms might be present and (*b*) learning about inherently endogenous causal processes requires us to study actions and decisions jointly, not separately, and this in turn requires approaching them by combining multiple methodologies and epistemologies. These tools already exist; what we need is to learn to view them as complements rather than rivals, and to develop common conceptual frameworks and terminology to facilitate constructive engagement across them. This review aims to provide some initial steps toward this, but fully exploring the implications and possibilities represents a much broader avenue for researchers across political science and adjacent disciplines.

One tangible recommendation for all researchers is to be more explicit in defining the scope of the causal systems they study. Do their theoretical and empirical frameworks include or exclude the upstream decisions that determine the independent variables they examine? What are the potential connections of these upstream processes to the downstream effects of these actions? What are the implications of these research design choices for how we interpret and apply the results?

Thinking about the role of decisions and agency within causal systems can also push us to be more explicit about the type of answers we want and thus the type of questions we ask. Levi Martin's (2011) distinction between the third- and first-person styles of explanation is useful in this regard. Whereas third-person perspectives on causality focus on uncovering abstract, generalizable causal relationships that exist regardless of (perhaps even despite) the agency of any individual(s), first-person perspectives center individual cognition and agency in understanding the causes and consequences of actions. These different perspectives imply different explanations and different questions and prompt us to ask whether and when our goal as scholars is to provide convincing explanations (third-person) or guides to more effective action (first-person). Are we interested in understanding the conditions under which social movements succeed at making change or communities organize to manage their natural resources sustainably, or are we interested in helping actors understand how to reform their bureaucracies? Both types of question are important and valid for different purposes and audiences, of course. But what is unique about inherently endogenous causal processes is that understanding them requires us to be able to simultaneously make space for both first- and third-person notions of causality within our theoretical frameworks and empirical analyses. Finding ways to do so might help us all produce research that not only provides more complete explanations but also is more relevant to the most important phenomena, challenges, and decisions confronting the world.

DISCLOSURE STATEMENT

ACKNOWLEDGMENTS

It would not exist without the challenging and fertile intellectual environment provided by the Center for Advanced Study in the Behavioral Sciences (CASBS) at Stanford University, and I am especially grateful to all the staff and members of the 2022–23 fellowship cohort there. All remaining mistakes are my own.

LITERATURE CITED

Abbott A. 2005. Linked ecologies: states and universities as environments for professions. *Sociol. Theory* 23(3):245–74

Abimbola S. 2018. On the meaning of global health and the role of global health journals. *Int. Health* 10:63–65

Abimbola S. 2019. The foreign gaze: authorship in academic global health. *BMJ Glob. Health* 4(5):e002068

Acemoglu D. 2010. Theory, general equilibrium, and political economy in development economics. *J. Econ. Perspect.* 24(3):17–32

Adler E, Pouliot V, eds. 2011. *International Practices*. Cambridge University Press

Agrawal A, Erbaugh J, Pradhan N. 2023. The commons. *Annu. Rev. Environ. Resour.* 48:531–58

Allcott H. 2015. Site selection bias in program evaluation. *Q. J. Econ.* 13(3):1117–65

Al-Ubaydli O, List JA, LoRe D, Suskind D. 2017. Scaling for economists: lessons from the non-adherence problem in the medical literature. *J. Econ. Perspect.* 31(4):125–44

Ansari SM, Fiss PC, Zajac EJ. 2010. Made to fit: how practices vary as they diffuse. *Acad. Manag. Rev.* 35(1):67–92

Aronow PM, Carnegie A. 2013. Beyond LATE: estimation of the average treatment effect with an instrumental variable. *Political Anal.* 21(4):492–506

Bates RH, Greif A, Levi M, Rosenthal J-L, Weingast B. 1998. *Analytic Narratives*. Princeton University Press

Bold T, Kimenyi M, Mwabu G, Ng'ang'a A, Sandefur J. 2018. Experimental evidence on scaling up education reforms in Kenya. *J. Public Econ.* 168:1–20

Bueno de Mesquita E, Berry CR, Ashworth S. 2021. *Theory and Credibility*. Princeton University Press

Bueno de Mesquita E, Tyson SA. 2020. The commensurability problem: conceptual difficulties in estimating the effect of behavior on behavior. *Am. Political Sci. Rev.* 114(2):375–91

Capoccia G. 2015. Critical junctures and institutional change. In *Advances in Comparative-Historical Analysis*, ed. J Mahoney, K Thelen. Cambridge University Press

Carpenter D. 2024. Just how much of history is countable? *Political Sci. Q.* 139(4):599–610

Cason TN, Tabarrok A, Zubrickas R. 2025. Signaling quality: how refund bonuses can overcome information asymmetries in crowdfunding. *Manag. Sci.* 71(7):5933–47

Chassang S. 2010. Building routines: learning, cooperation, and the dynamics of incomplete relational contracts. *Am. Econ. Rev.* 100(1):448–65

Clemens MA, Tiongson ER. 2017. Split decisions: household finance when a policy discontinuity allocates overseas work. *Rev. Econ. Stat.* 99(3):531–43

Coburn CE. 2016. What's policy got to do with it? How the structure-agency debate can illuminate policy implementation. *Am. J. Educ.* 122:465–75

Dal Bó P, Foster A, Putterman L. 2010. Institutions and behavior: experimental evidence on the effects of democracy. *Am. Econ. Rev.* 100:2205–29

Deaton A. 2010. Instruments, randomization, and learning about development. *J. Econ. Lit.* 48:424–55

Donaldson ML, Mavrogordato M, Youngs P, Dougherty S, Al Ghanem R. 2021. "Doing the 'real' work": how superintendents' sensemaking shapes principal evaluation policies and practices in school districts. *AERA Open* 7(1):2332858420986177

Fearon JD. 1995. Rationalist explanations for war. *Int. Organ.* 49(3):379–414

Findley MG, Kikuta K, Denly M. 2021. External validity. *Annu. Rev. Political Sci.* 24:265–93

Finnemore M, Wendt A. 2024. When "old" constructivism was new: reflections on classical constructivism. *Perspect. Politics* 22(4):1248–68

Gartzke EA. 1999. War is in the error term. *Int. Organ.* 53(3):567–87

Gartzke EA, Poast P. 2017. Empirically assessing the bargaining theory of war: potential and challenges. *Oxford Research Encyclopedia of Politics*. **https://doi.org/10.1093/acrefore/9780190228637.013.274**

Gechter M, Hirano K, Lee J, Mahmud M, Mondal O, et al. 2023. Site selection for external validity: theory and an application to mobile money in South Asia. Preprint, arXiv:2405.13241v1 [econ.GN]

George AL, Bennett A. 2005. *Case Studies and Theory Development in the Social Sciences*. MIT Press

Gibbons R, Kaplan RS. 2015. Formal measures in informal management: Can a balanced scorecard change a culture? *Am. Econ. Rev.* 105(5):447–51

Gibbons R, Matouschek N, Roberts J. 2013. Decisions in organizations. In *The Handbook of Organizational Economics*, ed. R Gibbons, J Roberts. Princeton University Press

Gilson L, ed. 2012. *Health Policy and Systems Research: A Methodology Reader*. World Health Organization

Glynn AN, Ichino N. 2015. Using qualitative information to improve causal inference. *Am. J. Political Sci.* 59(4):1055–71

Goertz G. 2017. *Multimethod Research, Causal Mechanisms, and Case Studies: An Integrated Approach*. Princeton University Press

Hawe P. 2015. Lessons from complex interventions to improve health. *Annu. Rev. Public Health* 36:307–23

Heckman JJ, Urzua S. 2010. Comparing IV with structural models: what simple IV can and cannot identify. *J. Econ.* 156(1):27–37

Hedström P, Ylikoski P. 2010. Causal mechanisms in the social sciences. *Annu. Rev. Sociol.* 36:49–67

Humphreys M, Jacobs A. 2023. *Integrated Inferences*. Cambridge University Press

Imai K, Keele L, Tingley D, Yamamoto T. 2015. Unpacking the black box of causality: learning about causal mechanisms from experimental and observational studies. *Am. Political Sci. Rev.* 105(4):765–89

Imbens GW. 2010. Better LATE than nothing: some comments on Deaton (2009) and Heckman and Urzua (2009). *J. Econ. Lit.* 48:399–423

Kentikelenis AE, Babb S. 2019. The making of neoliberal globalization: norm substitution and the politics of clandestine institutional change. *Am. J. Sociol.* 124(6):1720–62

Killick A. 2023. *Politicians and Economic Experts: The Limits of Technocracy*. Columbia University Press

Kowalski A. 2023. How to examine external validity within an experiment. *J. Econ. Manag. Strategy* 32(3):491–509

Leaver C, Ozier O, Serneels P, Zeitlin A. 2021. Recruitment, effort, and retention effects of performance contracts for civil servants: experimental evidence from Rwandan primary schools. *Am. Econ. Rev.* 111(7):2213–46

Levi M. 2006. Why we need a new theory of government. *Perspect. Politics* 4(1):5–19

Levi Martin J. 2011. *The Explanation of Social Action*. Oxford University Press

Lieberman ES. 2005. Nested analysis as a mixed-method strategy for comparative research. *Am. Political Sci. Rev.* 99(3):435–52

Lowande K, Salis-Muñiz I. 2025. When presidents limit bureaucratic power: evidence from abortion bans in foreign aid. *Perspect. Politics* 23(2):649–64

Mahoney J. 2010. After KKV: the new methodology of qualitative research. *World Politics* 62(1):120–47

Mahoney J, Thelen K, eds. 2015. *Advances in Comparative-Historical Analysis*. Cambridge University Press

Mangla A. 2022. *Making Bureaucracy Work: Norms, Education and Public Service Delivery in Rural India*. Cambridge University Press

Mansoor Z, Williams MJ. 2024. Systems approaches to public service delivery: methods and frameworks. *J. Public Policy* 44(2):258–83

Mansuri G, Rao V. 2013. *Localizing development: Does participation work?* Policy Res. Rep., World Bank

Marchal B, van Belle S, van Olmen J, Hoerée T, Kegels G. 2012. Is realist evaluation keeping its promise? A review of published empirical studies in the field of health systems research. *Evaluation* 18(2):192–212

Morrow JD. 1989. Capabilities, uncertainty, and resolve: a limited information model of crisis bargaining. *Am. J. Political Sci.* 33:941–72

Ostrom E. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge University Press

Padgett JF, Powell WW, eds. 2012. *The Emergence of Organizations and Markets*. Princeton University Press

Paul LA, Healy K. 2018. Transformative treatments. *Noûs* 52(2):320–35

Pettigrew AM. 1990. Longitudinal field research on change: theory and practice. *Organ. Sci.* 1(3):267–92

Piotrowski SJ, Berliner D, Ingrims A. 2022. *The Power of Partnership in Open Government: Reconsidering Multistakeholder Governance Reform*. MIT Press

Popp Berman E. 2022. *Thinking Like an Economist: How Efficiency Replaced Equality in U.S. Public Policy*. Princeton University Press

Ragin CC. 2014. *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies*. University of California Press

Repucci S. 2014. Designing effective civil service reform lessons from past experience. *Public Adm. Dev.* 34:207–18

Riker WH. 1984. The heresthetics of constitution-making: the presidency in 1787, with comments on determinism and rational choice. *Am. Political Sci. Rev.* 78(1):1–16

Sandberg J, Tsoukas H. 2015. Making sense of the sensemaking perspective: its constituents, limitations, and opportunities for further development. *J. Organ. Behav.* 36:S6–32

Schatz E, ed. 2009. *Political Ethnography: What Immersion Contributes to the Study of Power*. University of Chicago Press

Shipan CR, Volden C. 2008. The Mechanisms of Policy Diffusion. *Am. J. Political Sci.* 52(4):840–57

Slough T, Rubenson D, Levy R, Rodriguez FA, del Carpio MB, et al. 2021. Adoption of community monitoring improves common pool resource management across contexts. *PNAS* 118(29):e2015367118

Slough T, Tyson SA. 2023. External validity and meta-analysis. *Am. J. Political Sci.* 67(2):440–55

Slough T, Tyson SA. 2024. *External Validity and Evidence Accumulation*. Cambridge University Press

Stevenson M. 2023. Cause, effect, and the structure of the social world. *Boston Univ. Law Rev.* 103(7):2001–47

Thelen K, Mahoney J. 2015. Comparative-historical analysis in contemporary political science. In *Advances in Comparative-Historical Analysis*, ed. J Mahoney, K Thelen. Cambridge University Press

Thiel A, Blomquist WA, Garrick DE. 2019. *Governing Complexity: Analyzing and Applying Polycentricity*. Cambridge University Press

Tolbert PS, Zucker LG. 1983. Institutional sources of change in the formal structure of organizations: the diffusion of civil service reform, 1880–1935. *Adm. Sci. Q.* 28(1):22–39

Tsikata YM. 2001. Owning economic reforms: a comparative study of Ghana and Tanzania. WIDER Discuss. Pap. 2001/53, The United Nations University World Institute for Development Economics Research

Tsoukas H, Chia R. 2002. On organizational becoming: rethinking organizational change. *Organ. Sci.* 13(5):567–82

von Soest C. 2023. Why do we speak to experts? Reviving the strength of the expert interview method. *Perspect. Politics* 21(1):277–87

Wawro GJ, Katznelson I. 2022. *Time Counts: Quantitative Analysis for Historical Social Science*. Princeton University Press

Weick KE. 1995. *Sensemaking in Organizations*. Sage

Widner J, Woolcock M, Nieto DO. 2022. *The Case for Case Studies: Methods and Applications in International Development*. Cambridge University Press

Williams MJ. 2020. External validity and policy adaptation: from impact evaluation to policy design. *World Bank Res. Obs.* 35(2):158–91

Williams MJ. 2026. *Reform as Process: Implementing Change in Public Bureaucracies*. Columbia University Press

Yanow D, Schwartz-Shea P, eds. 2014. *Interpretation and Method: Empirical Research Methods and the Interpretive Turn*. Routledge. 2nd ed.