The Credibility Revolution and the Empirical Implications of Theoretical Models

Tara Slough*

April 11, 2024

Abstract

This chapter connects two important developments in methodology in the past twenty years: the Empirical Implications of Theoretical Models (EITM) project and the credibility (or identification) revolution. I argue that the emphasis on aggregate causal estimands (quantities of interest) advanced by the credibility revolution *can* be used to advance the EITM project's overarching goal to connect theoretical models and quantitative empirical analysis. In an effort to facilitate broader integration of applied theory with credibility-motivated research designs, present a new framework for existing and future efforts to map applied theory to (reduced-form) causal estimands. The framework is organized around three properties of these estimands: (1) the definition of treatment effects within the potential outcomes framework; (2) the aggregation of individual treatment effects; and (3) the determination of the set of units over which this aggregation occurs. This organization reveals a need for further analyses that links theory to considerations of aggregation (2) and subgroup selection (3).

*Assistant Professor, New York University. tara.slough@nyu.edu

This volume reflects upon twenty years of the Empirical Implications of Theoretical Models (EITM) project. EITM originated from a National Science Foundation (NSF) workshop that sought to understand and remedy a "schism" between political scientists engaged in formal modeling and those engaged in quantitative empirical analysis (Political Science Program, Directorate for Social, Behavioral and Economic Sciences, 2001). In this 2001 workshop, the NSF gathered a group of leading political scientists and scholars in cognate fields to comment upon this divide and proposed efforts to bridge this gap.

In the course of those same twenty years, empirical analysis in political science has undergone a large-scale transformation, often called the "credibility revolution" or the "identification revolution" (Leamer, 1983; Angrist and Pischke, 2010; Samii, 2016).[1] The credibility revolution promotes the use of research designs that help to facilitate identification and estimation of various *aggregate causal effects*. I use the term "aggregate causal effects" because these estimands provide summaries of the causal effects across multiple units. These estimands are, in general, distinct from the regression coefficients and marginal effects that (implicitly) served as estimands in empirical political science at the time of the original EITM report (Political Science Program, Directorate for Social, Behavioral and Economic Sciences, 2001).

A prominent argument of critics of the credibility revolution holds that the adoption of credibility-inspired research designs furthered the chasm between applied theory and quantitative empirical work in political science by de-emphasizing the role of theory (Clark and Golder, 2015; Samii, 2016; Huber, 2017; Franzese, 2020).[2] In contrast to these views, I argue that the credibility revolution *facilitates* our ability to connect applied theory to empirical research. In these cases, the central benefit of the credibility revolution is its clarification of empirical estimands. When these estimands are clearly specified, it easier to translate (formally) the empirical targets produced by a theory and the estimands targeted by a research design. Such analysis can be used identify

[1]The credibility revolution is not unique to political science. It has also proceeded in other social scientific disciplines over a similar period, albeit at somewhat different rates.

[2]But see Ashworth, Berry, and Bueno de Mesquita (2015) and Ashworth, Berry, and de Mesquita (2021) for discussion of the importance of "all else equal" claims in both credibility-oriented empirical research and in many social scientific theories.

problems of commensurability between the theory and research design: situations in which the empirical target, or quantity of theoretical interest, does not align with the estimand (Bueno de Mesquita and Tyson, 2020; Ashworth, Berry, and de Mesquita, 2021).

Analyses that link applied theory to various aggregate causal effects have identified commensurability problems of different forms (e.g., Eggers, 2017; Bueno de Mesquita and Tyson, 2020; Slough, 2023; Abramson, Koçak, and Magazinnik, 2022). These findings may bolster critics' fears of a chasm between theory and empirics using credibility-inspired designs. However, they do not provide support for claims of a widening chasm *relative to earlier approaches*. Indeed, analyses of the type undertaken by the these authors would be much more difficult/less tractable if the quantities of interest were more stylized or bespoke to a specific statistical model, as was the case for concern among the members of the original EITM workshop participants. Thus, one advance of the credibility revolution that furthers the scope of EITM is that it allows us to more clearly articulate the relationship between (reduced-form) empirical estimands and the empirical targets produced by a (formal) theory.

While scholars have begun to identify a variety of commensurability problems in various applied literatures, we lack a general framework for organizing these analyses. This obscures our understanding of the relationship how various commensurability problems that have already been identified relate to each other. It further limits our conception of how widespread or important these problems might be in applied empirical research. To develop such a framework, one could, in principle, adopt one of two approaches. Specifically a framework could be organized around features of empirical targets or characteristics of aggregate causal effects.

Relative to a framework organized around empirical targets, a framework organized around properties of aggregate causal effects is more useful for connecting existing and future commensurability analyses. There are many approaches to theory in the social sciences. A framework organized around empirical targets would require some broad structural similarities across these varied approaches that are relevant for linking theory to causal estimands.[3] Identifying such struc-

[3]Bueno de Mesquita and Tyson (2020) and Slough (2023) provide examples of this approach. While both works are relevant to a substantial number of relevant empirical research designs,

tural similarities is bound to be a challenging endeavor. In contrast, the aggregate causal effects privileged by credibility revolution practitioners adopt a highly standardized structure. This structure proves quite useful for organizing existing critiques and generating new ideas.

In this chapter, I propose a framework for understanding the relationship between commensurability critiques premised on three common features of aggregate causal effects. First, I consider the use of the *potential outcomes framework* to define causal effects. A majority of existing commensurability critiques identify problems in the mappings between theoretical concepts of interest and potential outcomes that researchers seek to measure. Second, commensurability issues routinely arise in the *aggregation* of causal effects across subjects. While these issues are explored in the context of aggregating preferences through conjoint or multidimensional choice experiments (Abramson, Koçak, and Magazinnik, 2022), they emerge in effectively all causal identification-oriented research designed. Finally, and least explored, are commensurability considerations related to *which units* to aggregate over when estimating an aggregate causal effect. This framework allows for new connections between existing commensurability critiques. It also suggests a need for further consideration of aggregation and selection of (subsets of) units in future work.

I proceed by providing background on the relationship between credibility revolution-based empirical approaches and the multivariate regression models that most concerned the original EITM working group. This comparison suggests an important standardization of empirical quantities of interest. Drawing on common features of these quantities, I organize existing analyses of commensurability with an eye to where additional analysis is needed. I conclude with suggestions for how this agenda can be designed to forge tighter connections between reduced-form empirical strategies advanced by the credibility revolution and applied theory.

1 The Credibility Revolution

The credibility revolution refers to an intellectual movement in the empirical social sciences toward the use of research designs that facilitate—or make more credible—claims for the identification

neither is sufficiently broad to organize all existing commensurability critiques, let alone those that are yet to be articulated.

of causal effects (Samii, 2016). Angrist and Pischke (2010) trace the origins of this movement in economics to Leamer's (1983: p. 37) exhortation to "take the con out of econometrics." While the motivations and sequencing of this revolution across different social science disciplines vary somewhat, the philosophical commitments of the broader movement are remarkably similar. Specifically, Slough and Tyson (2024: p. 3) summarize these commitments as follows:

1. *A model of causality*: Causality is defined within the potential outcomes [framework];
2. *Methodological commitments*: Identification arises from a model of a research design rather than a model of the data-generating process;
3. *Evaluating estimators*: Prioritization of unbiasedness over other properties of estimators.

Efforts to connect applied theory to crediblility-driven research designs relate directly to commitments #1 and #2. Articulation of a theory provides guidance on how potential outcomes—and thus treatment effects—are defined (e.g., Slough, 2023). Theory further sheds light on what effect is identified given a counterfactual comparison within a research design (e.g., Bueno de Mesquita and Tyson, 2020). By boiling the credibility revolution down to a set of core commitments, the importance of applied theory for defining, identifying, and interpreting causal effects within this intellectual tradition becomes clear. It is useful to emphasize that efforts to connect theory to causal effects in this way reveal a need to consider methodological challenges that are distinct from concerns about estimation.

1.1 Causal Estimands

A primary change in empirical practice associated with the credibility revolution has been a clarification of the *estimand*, or the quantity to be estimated, in applied research. Credibility-motivated research designs—including experimental, regression discontinuity, difference-in-differences, and instrumental variables designs—target a set of *aggregate causal effects*, typically defined or mo-

tivated in terms of differences or changes in potential outcomes (Neyman, 1923; Rubin, 1974).[4] Potential outcomes describe the set of outcomes that *would* manifest under different realizations of treatment. Of course, at the unit level, we observe only one realized potential outcome, meaning that we do not observe unit-level causal effect. As a consequence, standard causal estimands represent a quantitative summary of differences in potential outcomes across multiple units.

To illustrate the general structure of these estimands, denote potential outcomes for outcome Y, under treatment or instrument level ω, as $Y_i(\omega)$.[5] Causal estimands are defined at the aggregate level, i.e., across multiple units (indexed by i). The operator $f(\cdot)$—typically the expectations operator or a quantile function—provides a measure of an aggregate treatment effect. Some estimands further condition the set of units over which the treatment effect is evaluated. For example, a regression-discontinuity design estimates the local average treatment effect (LATE) at a threshold in the running variable. A difference-in-differences design with a single post-treatment period estimates the average treatment effect on the treated (ATT), which is the ATE among treated units. Formally, we can say that the aggregate treatment effect is evaluated for a subset of units, $i \in \mathcal{D}$, where the set $\mathcal{D}$ depends on the design and estimand. This minimal formalization provides a straightforward characterization of causal effects, as follows:

$$f_{\mathcal{D}}(Y_i(\omega'') - Y_i(\omega')) \tag{1}$$

The estimand in (1) is sufficiently general to characterize the vast majority of causal effects that are estimated in much of the design-based literature, with two exceptions. First, instrumental variable designs typically seek to assess the (local) effect of an endogenous treatment rather than the exogenous instrument ω. However, it is straightforward to extend the estimand in (1) to accomodate this case.[6] Second, in some cases, instruments take more than values (or dosages), such

[4] Adherents to Pearl sometimes express causal estimands in terms of the do-calculus.

[5] I adopt the notation used in Slough and Tyson (2023) and Slough and Tyson (2024) to characterize estimands/empirical targets.

[6] Specifically, define $D(\omega)$ as the endogenous treatment. For the binary instrument and binary treatment case, the LATE among compliers is given by $E_{\mathcal{D}}[Y_i(D'') - Y_i(D')]$, where $\mathcal{D}$ is the

that we are interested in characterizing how an outcome responds to more than two values of an instrument (i.e., tracing a dosage response curve). The concerns that I describe in the context of binary treatments transport to settings with a multi-valued instrument. I focus on the binary case in the interest of parsimony.

1.2 Relationship to Multivariate Regression

Our discussion of causal estimands focuses on the quantities that empirical researchers seek to estimate using various research designs. To this point, I have not discussed how these quantities are estimated. Indeed, for many such designs, researchers use multivariate regression to estimate these quantities of interest. In this sense, examination of a regression table used to report, for example, estimates of the ATE from a randomized experiment, may look indistinguishable from a regression tables characteristic of much quantitative empirical work from *before* the credibility revolution. In contrast to earlier multivariate approaches, however, credibility-based approaches: (1) clarify the causal estimand; and (2) devote more attention to avoiding biased estimators of these quantities.

In the absence of a clear estimand, earlier multivariate approaches typically emphasized estimation of regression coefficients. While regression coefficients can be *estimators* of causal estimands, as above, they are not, in general, interpretable as such. Moreover, even for purposes of description—as opposed to causal identification—the coefficient from a multivariate regression is often several steps removed from any underlying quantity of interest (e.g., a conditional association between two variables). Lundberg, Johnson, and Stewart (2021) argue that treating regression coefficients as estimands quickly devolves into tautology, writing (p. 533, emphasis original):

> "That mode of inquiry defines the research goal *inside* a particular statistical model. If your research goal is a coefficient of a particular model, then you are committed to that model: it becomes impossible to reason about other approaches to achieve the goal. By contrast, we advocate a statement of the goal *outside* the statistical model—like

stratum of compliers, those units for whom $D(\omega'') = D''$ and $D(\omega') = D'$.

> an average causal effect or a population mean——which opens the door to alternative estimation procedures that could answer the research question under more credible assumptions."

The original EITM working group was concerned primarily with multivariate regression analysis and many recognized these issues. In particular, they noted that regression coefficients were too often treated as estimands of theoretical interest. For example, Henry Brady commented "there are still far too many statistical 'modelers' who think that a regression equation or a likelihood function constitutes a model" (Political Science Program, Directorate for Social, Behavioral and Economic Sciences, 2001: p. 37). Rebecca Morton noted "most political science began to believe that statistical models were theoretical models and that just writing a regression equation. . .was theorizing" (Political Science Program, Directorate for Social, Behavioral and Economic Sciences, 2001: p. 54). In other words, these authors comment on the perils of confusing a *statistical* model for a *theoretical* model.

The confusion between statistical and theoretical models has arguably been reduced as a function of the adoption of credibility-inspired research designs. Specifically, the emphasis on causal estimands, rather than coefficients of a regression model, generally simplifies the process of linking theory to empirical estimands. Nevertheless, a number of authors studying different identification-oriented research designs and applications point out that this link from theory to estimand is often implicit in applied research (Bueno de Mesquita and Tyson, 2020; Ashworth, Berry, and de Mesquita, 2021; Slough, 2023). The lack of explicit probing of these links raises the possibility that theoretical quantities of interest—termed *empirical targets* by Slough and Tyson (2023)—do not align with the estimands identified by a research design.

The credibility revolution obviously does not guarantee commensurability of empirical targets and causal estimands produced by a research design. However, because these estimands are defined *outside* of a statistical model, it is substantially easier to construct or probe mappings between empirical targets and estimands. This represents a central contribution of the credibility revolution to further the goal of the EITM agenda.

1.3 EITM and the Theoretical Implications of Empirical Models

Many recent works that examine connections between credibility-inspired research designs and applied theory describe their approach as the "theoretical implications of empirical models" (TIEM) (for introduction of this term, see Wolton, 2019). TIEM represents a broadening of the scope of ways that theory and empirical work can be integrated. EITM scholarship has traditionally emphasized use of empirical tests to evaluate comparative static predictions of theoretic models. In contrast, work classified as TIEM expands the connection between theory and empirics to include considerations of mechanisms, theoretical barriers to identification of a treatment effect, interpretation of measured effects, and accumulation of evidence (Wolton, 2024; Slough and Tyson, 2024).

To make progress on these questions, TIEM work takes the step of incorporating explicitly features of empirical research design into formal models of substantive phenomena (e.g., Ashworth, Berry, and Bueno de Mesquita, 2024; Slough and Tyson, 2024). Such analyses are facilitated by the clear specification of causal effects as quantities of interest in credibility-motivated empirical work. Whether TIEM should be viewed as a part of EITM, a generalization of EITM, or a distinct intellectual movement remains an open question that is beyond the scope of this chapter. In the remainder of this chapter, I provide a framework for organizing existing works described as TIEM and identifying topics that merit more consideration.

2 A Framework Organized around Causal Estimands

Providing links between empirical targets from a theory and the estimands produced by a research design helps to clarify: (1) what theoretical quantity is being estimated; and (2) what substantive inferences can be drawn from a given empirical exercise. But how much generality is possible when we develop links between theory and empirical research design? For a framework to be broadly applicable, we must be able to abstract beyond a specific class of research designs or theoretical models. Understanding the scope for findings about links between empirical targets and causal estimands is important because it clarifies the nature of EITM efforts. If we ultimately need a theoretical model for every research design-treatment-outcome combination, there is arguably less

justification for general consumption or use of this work. Yet, efforts identify structural similarities across different models or research designs, analyses that identify or solve issues related to the mapping from theories to causal estimands should, in theory, have more impact. Impact may come in the form of broader awareness of an issue or new suggestions for research design.

There are different ways to identify structural similarities across theories and/or research designs. For example, in other work (Slough, 2023), I propose a general classification of dynamic theoretical models—strategy set symmetry—that is relevant for identification of causal estimands. In this chapter, I begin instead from features of the class of causal estimands given by (1). I identify three features of these estimands that pose interesting implications from the mapping from empirical targets to estimands:

1. Defining potential outcomes and treatment effects: $Y_i(\omega'') - Y_i(\omega')$

2. Aggregating treatment effects: $f(\cdot)$

3. Local and conditional effects: defining the set $\mathcal{D}$

In the remainder of this chapter, I discuss what we know about these three features of causal estimands and their relationship to applied theory. The goal of this discussion is to provide a framework for organizing existing links between .

3 Defining potential outcomes and treatment effects

Credibility revolution-inspired work typically defines causal effects within the potential outcomes framework or Neyman-Rubin causal model (Rubin, 1974; Holland, 1986). The potential outcomes framework serves as a model of the world in which a potential outcome is expressed as a function of some treatment instrument (ω) and some outcome $Y_i(\omega)$ for a unit, i. In most applied empirical research researchers multiple multiple outcomes, e.g., $Y_i^k(\omega)$ where $k \in \{1, ..., K\}$ indexes distinct outcomes. In conventional presentations of the potential outcomes framework, the relationships between different outcomes are unspecified. While outcomes are ostensibly related via dependence on an instrument, ω, analysis of treatment effects on different outcomes is typically conducted

independently (without reference to other outcomes) and symmetrically (e.g., estimating the ATE on each of the K outcomes).[7]

3.1 Multiple Outcomes

Applied theory typically imposes more structure on the relationships between outcomes of interest than standard expositions of the potential outcomes framework. For example, consider an information and electoral accountability experiment. An experimenter exogenously provides (or amplifies) an informational signal about an incumbent's action or performance in the electorate. They then measure subjects' beliefs about the incumbent's type and subjects' (self-reported) vote choices. This design yields two outcomes of interest: a measure of beliefs (Y_i^1) and vote choice (Y_i^2). Under the potential outcomes framework, one could estimate following treatment effects (holding fixed f and $\mathcal{D}$ for the sake of clarity):

$$f_{\mathcal{D}}(Y_i^1(\omega'') - Y_i^1(\omega')): \quad \text{the effect of } \omega'' \text{ relative to } \omega' \text{ on beliefs}$$

$$f_{\mathcal{D}}(Y_i^2(\omega'') - Y_i^2(\omega')): \quad \text{the effect of } \omega'' \text{ relative to } \omega' \text{ on vote choice.}$$

How does a theoretical model discipline the relationship between potential outcomes $Y_i^1(\omega)$ and $Y_i^2(\omega)$? It posits how the two outcomes—beliefs and vote choice—are related. Such a model of the experimental environment will typically proceed by suggesting that voters value an incumbent's type and other attributes of a candidate (e.g., distance between a voter's ideal point and the candidate's platform) or valence.[8] In the model, once citizens update their beliefs, they choose between candidates on the basis of this belief and other attributes of their preferences over candidates. This observation is helpful in understanding *differences* in the respective the effects of

[7] Two exceptions to this generalization are settings with noncompliance with treatment assignment and mediation analysis. In each case, asymmetric treatment of different outcomes is employed to estimate different causal effects (estimands) for different outcomes.

[8] Various models of information and accountability experiments and observational studies exist. See Izzo, Dewan, and Wolton (2020) or Slough and Tyson (2024) for examples.

information on beliefs about the incumbent and on vote choice, the two outcomes of interest. So long as voters value an incumbent's type (over which they hold beliefs), these outcomes should be related. Measured treatment effects however *should be different* between the two outcomes for two reasons. First, the outcomes are defined on different scales. A belief about the incumbent's type is, in principle, a (continuous) probability whereas vote choice is a discrete outcome under all standard electoral systems. Second, vote choice is a function of voter beliefs about the incumbent, and is therefore causally subsequent to voter beliefs.[9]

The two features of outcomes described in the above example—causal sequencing and range of different outcomes—are important for identification, interpretation, and estimation. Applied theory can guide thinking on these two fronts. First, not all outcomes that occur after treatment are defined under a given theory. In Slough (2023), I argue that in settings with sequential behavioral outcomes, a game tree—even one without utililites or an information structure—is informative about which (potential) outcomes are defined. When a game or decision tree is asymmetric, meaning that at some non-terminal node, the player or set of actions availability to that player is different, a tree exhibits *strategy set asymmetry*. Behavioral outcomes that are (theoretically) realized after the first strategy set asymmetric node are undefined for some histories of the game. If potential outcomes are undefined, so too are aggregate causal estimands that aggregate outcomes. This is analogous to the truncation by death problem in medicine (Frangakis and Rubin, 2002; McConnell, Stuart, and Devaney, 2008). In this case, applied theory is helpful for disciplining about which potential outcomes are defined for all units in a sample or population under a given conceptual framework.

Second, the sequencing and range of potential outcomes poses under-recognized challenges for interpretation of causal effects. For example, in Fu and Slough (2024), Jiawei Fu and I examine when heterogeneous treatment effects (HTEs) or comparison of subgroup treatment effects are informative about mechanism activation. More specifically, we ask when the existence of HTEs suggests that a mechanism is activated for at least one unit in the sample. The challenge, as we

[9]Indeed, voter beliefs could be viewed as mediator relative to the vote choice outcome.

show, is that the use of HTEs for mechanism activation relies on the additive separability of the indirect effect of a mechanism from the effects of other mechanisms. When a mechanism's effects on an outcome are then mapped by some non-linear function into a subsequent outcome—as is the case with beliefs and (discrete) vote choice in our example—this non-linear transformation breaks the additive separability of the learning mechanism from other potential mechanisms. The existence of HTE with respect to some pre-treatment covariate on vote choice therefore cannot support a claim about mechanism activation.

In empirical work, substantial effort has been devoted clarifying issues of timing relative to treatment. These concerns largely revolve around the threat of post-treatment bias (see, for example, Montgomery, Nyhan, and Torres, 2018). Issues related to relationships between multiple outcomes are less explored. In contrast, applied theory is more attentive to the structure of a causal process and the set of outcomes produced. By analyzing the relationship between multiple outcomes theoretically, researchers can gain insights about both the identification and interpretation of treatment effects on these outcomes.

3.2 Selecting Instruments

A potential outcome is defined relative to a specific instrument. By asserting that an outcome, Y_i, is a function of an instrument, ω', we expect that the potential outcome could change if we were to employ a different instrument $\omega'' \neq \omega'$. It is conventional to give labels (e.g., "treatment" and "control") or numerical normalizations (e.g., "1" and "0") to the instruments employed in credibility-driven empirical studies. But connecting these measures to underlying concepts is essential to understanding what a measured treatment effect might convey about substantive phenomena of interest. To this end, applied theory is useful for helping researchers to articulate *how* a specific contrast of two or more instruments could induce a change in a potential outcome. These considerations are important for understanding what a treatment effect is measuring. A large share of TIEM work focuses on this important task of interpreting what can be inferred about a phenomenon or mechanism from a given contrast.

Bueno de Mesquita and Tyson (2020) consider the case of causal studies that seek to estimate the effect of an actor A's behavior on the behavior of another actor B. In many contexts, actor A's behavior could affect actor B's behavior both through a direct channel and an informational channel and researchers would like to measure the total effect of these two channels. In a motivating example of Bueno de Mesquita and Tyson (2020), many political scientists are interested in studying the effect of protest on government behavior. The total effect of protest on government behavior comes from two distinct effects: a direct effect in which a government responds to demands with repression and/or concessions and an (indirect) informational effect in which the government learns about the underlying level of grievance in the population, which could lead to changes in the use of repression/concessions.

Since protests are not random events, credibility-motivated researchers are likely to seek "as-if random" instruments that induce variation in protest size/occurrence in order to measure the effect of protest. For example, rainy days may suppress protest participation, so researchers contrast rainy to non-rainy days to assess the (first-stage) effect on protest, and then downstream effects on government responses to protest (e.g., Madestam et al., 2013). However, the government can also observe weather, which should generally affect what governments can learn from observing protest. Thus, the contrast of rainy to non-rainy days will yield a different informational effect than protests in general. This means the instrumental variables design cannot measure the total effect of protest. Moreover, Bueno de Mesquita and Tyson (2020) show that such a design can only isolate the direct effect of protests on behavior (absent the informational effect) unless: (1) there is no informational effect and (2) the effects of the behavioral and informational channels are additively separable with respect to the government behavior of interest.

Here, the potential lack of alignment between the empirical target produced by a theory and the estimand identified by the research design occurs because of the instruments chosen to causally identify an effect of protest. If researchers could practically or ethically randomly assign protest or find a valid instrument for protest that was not observable by the government, researchers could identify the total effect of protest on government response. As in this example, many examples

of commensurabiility problems in the literature stem from constraints in isolating (or creating) exogenous variation in instruments.

Applied theory is useful in clarifying what effect a contrast is identifying. Discussions of one popular research design for credibly estimating the causal effect of election outcomes—electoral regression discontinuity designs—illustrate the importance of this sort of theoretical analysis. Using this design, researchers seek to estimate the local average treatment effect of some election outcome in a tied election. In one popular application, researchers aim to study the effects of electing a specific type of candidate, e.g., a woman, on some subsequent outcome, e.g., legislative performance. Marshall (2022) argues that when candidate type predicts electoral success, marginal winners of different types are likely to vary on multiple dimensions. For example, if voters discriminate against female candidates, a female candidate who just wins her race is more likely to have some other feature that voters like (e.g., expertise) to compensate than a just-winning male candidate. Thus the contrast in this design is: female and more expert (on average) versus male and less expert (on average). Thus, estimated treatment effects cannot be as the effect of gender in isolation. For a specific example, in the context of women's underrepresentation in politics, Ashworth, Berry, and Bueno de Mesquita (2024) show how this form of compound treatment activates multiple mechanisms (e.g., higher quality thresholds and candidate luck). As a result, it is not possible to distinguish between multiple theoretical accounts of women's underrepresentation using the effects estimated from an electoral regression discontinuity design.

The above examples illustrate how articulation of theory aids in clarifying what can be learned about a phenomenon or mechanism from a given contrast. First, by specifying the relationship between the instruments in a contrast and potential outcomes researchers can be more clear about what effect is captured by a causal estimand. For example, Marshall (2022) shows that many regression discontinuity designs that compare marginal winners with different characteristics identify the effect of a compound treatment of a bundle of characteristics that jointly determine a candidate's electoral success. Second, the other examples seek to understand when a contrast can isolate the effect of a mechanism (Ashworth, Berry, and Bueno de Mesquita, 2024; Bueno de Mesquita

and Tyson, 2020). Other excellent examples of this use of applied theory include Eggers (2017). Finally, theory can be used to understand when the effect identified by a specific contrast is informative about the effect of an "ideal" (or "real world") contrast of interest (Bueno de Mesquita and Tyson, 2020).

4 Aggregating causal effects

Empirically, the motivation for studying aggregate—rather than individual—causal effects is straightforward. Substituting the notation from (1), Holland (1986: p. 947) describes the *fundamental problem of causal inference*, writing that "it is impossible to observe the value of $[Y_i(\omega'')]$ and $[Y_i(\omega')]$ on the same unit and, therefore, it is impossible to observe the effect of $[\omega'']$ on $[i]$." In response to this impossibility, empirical researchers aggregate over units in order to measure some form of aggregate causal effects. A broad focus on *averages*, e.g., the average treatment effect in an experiment, is a matter of convention and straightforward estimation.

Yet, this focus on identification and estimation of aggregate effects can introduce contradictions between conceptual objects and measured effects. Some of these issues emerge as a function of how we aggregate within or over different strategic interactions, but the strategic setting is not necessary for issues related to aggregation to emerge. I consider first issues of aggregation which are explored in recent debate over the use of conjoint (or multidimensional factorial choice surveys). I then discuss how aggregation of individual treatment effects represents an important—and largely neglected—terrain for connecting theory and reduced-form empirics.

4.1 Aggregating Preferences in Conjoint Surveys

In two influential articles, Hainmueller and Hopkins (2014) and Hainmueller, Hopkins, and Yamamoto (2014) popularized the use of conjoint or multidimensional choice experiments in political science. In these survey experiments, respondents compare profiles consisting of multiple randomized attributes. For example, respondents may choose between (hypothetical) political candidates who are characterized by three attributes: party, gender, and occupation. By comparing respondent choices between or ratings of these randomized candidate profiles, researchers seek to infer voter

preferences over different candidate characteristics.

Hainmueller, Hopkins, and Yamamoto (2014) advocate the estimation of average marginal component effects (AMCEs) from conjoint data. They characterize the AMCE as "the marginal effect of attribute l averaged over the joint distribution of the remaining attributes" (Hainmueller, Hopkins, and Yamamoto, 2014: p. 10). In part due to the ease of estimation, the AMCE has become the dominant estimand used in the analysis of conjoint experiments in political science. As an effort to learn about preferences—for example, voter preferences for female versus male candidates—the AMCE is, at best, a summary of *aggregate* preferences.

Abramson, Koçak, and Magazinnik (2022) document the limits of the AMCE as an interpretable measure of respondent preferences. In conjoint experiments in which respondents choose one profile, researchers routinely estimate the AMCE of an attribute level—for example, "male" for the candidate gender attribute—on respondents' choice of a profile. Where this AMCE of male relative to female profiles is positive, many researchers had interpreted the positive AMCE as evidence that a majority of respondents prefer male to female candidates. Abramson, Koçak, and Magazinnik (2022) show that this interpretation of a positive AMCE is unwarranted. Specifically, the AMCE incorporates both the extensive and intensive margins of preferences for candidate gender (or any other attribute/level). This means that a positive AMCE for male candidates could be generated by a majority of respondents preferring the male profile *or* a small minority of respondents intensely preferring the male profile. Ultimately, these possibilities cannot be distinguished from an AMCE estimate.[10]

In an ideal world, researchers that seek to characterize the preferences of respondents (or an electorate) would be able to elicit an individual's preferences over a battery of candidate or profile attribute combinations. With such data in hand, they could then (1) assess the correlates of different preference profiles; or (2) generate statements of aggregate preferences by assessing (for

[10] Abramson, Koçak, and Magazinnik (2022) provide bounds on the extensive margin of support for a candidate, namely the share of respondents that may prefer one attribute level—e.g., male candidates relative to female candidates—given an AMCE estimate. Note, however, for modest AMCE estimates and complex designs with many attributes and levels, these bounds are typically too wide as to be informative about a majority preference.

example) the share of individuals who favor male candidates in a given all-else-equal contest. Conjoint experiments are useful precisely because it is infeasible and/or prohibitively costly to elicit individual preferences over a large set of candidate characteristics. But, when researchers estimate AMCEs with data from a conjoint experiment, they rely on aggregation over individuals and profiles to make some statement about aggregate preferences. This estimand is not commensurable with many of the theoretically-relevant empirical targets (e.g., what attribute level is preferred by a majority) that we might want to know.

4.2 Aggregation Issues beyond Conjoint Surveys

Beyond the specific case of AMCEs in conjoint surveys, too little attention has been devoted to issues of aggregation of individual treatment effects more generally. In credibility-driven research (and quantitative empirical research more generally), researchers estimate *aggregate* causal effects. In the context of prevailing norms, however, theories articulate predictions about the beliefs or behavior of individual actors. In the accountability example above, for example, theory offers predictions about the beliefs and voting behavior of individual voters (or a single decisive voter). When expressed in terms of a "treatment effect," then, theoretical predictions map first to *individual* treatment effects. Mathematically, it is straightforward to aggregate these ITEs into an aggreate treamtent effect, i.e., the ATE, to express the empirical target produced by a theory (Slough and Tyson, 2024). But we lose information in this process of aggregation.

To illustrate, suppose that we have a theory to suggest a contrast produces treatment effects on an outcome by activating one or both of two distinct mechanisms, 1 and 2. To simply exposition, assume that we can express the ITE for unit i as:

$$ITE_i = M_{1i} + M_{2i}, \tag{2}$$

where M_{1i} is the effect of mechanism 1 (if activated) for individual i and M_{2i} is the effect of mechanism 2 (if activated). Moreover, suppose that theory suggests that, when activated, the effects of these mechanisms are counterveiling: mechanism 1 increases an outcome (e.g., $M_{1i} \geq 0$) while

mechanism 2 decreases it (e.g., $M_{2i} \leq 0$). If we could observe the sign of ITE_i for all units, i, we could answer questions of the form:

- For unit i does mechanism 1 or 2 have a larger (in magnitude) influence?
- For what share of units does mechanism 1 have a larger influence? For what share of units does mechanism 2 have a larger influence?

However, we know that we cannot observe any ITE_i, forcing us to rely on an aggregate treatment effect. Consider first the ATE, or $\mathbb{E}[ITE_i]$. If we were to observe a positive ATE, we could surmise that mechanism 1 has a larger influence than mechanism 2 for at least one unit ($M_{1i} \geq |M_{2i}|$ for some i). If we were to observe a negative ATE, we could surmise that M_{2i} has a larger influence for at least one unit ($M_{1i} \leq |M_{2i}|$ for some i). Thus, while our ability to detect and measure mechanism influence through reduced-form causal effects would be limited even if we could observe ITEs, it is clearly limited further due to the fact that we observe only aggregate causal effects.

Clearly, there are different ways to aggregate treatment effects. With the example in (2), for example, suppose that we could show treatment effect heterogeneity such that a contrast produces a positive treatment effect on some units and a negative treatment effect on others. In this case, we could say that there exist some unit(s) for which the influence of mechanism 1 is stronger than the influence of mechanism 2 and there exist some unit(s) for which the influence of mechanism 2 is stronger than the influence of mechanism 1. Showing variation in the signs of quantile treatment effects (instead of the ATE) represents one way to provide evidence for this claim. This suggests that the method by which we aggregate treatment effects matters for our ability to use theory to interpret reduced-form causal effects.

This example is clearly stylized. Because empiricists working in the credibility revolution tradition estimate *aggregate* causal effects, care must be given to how our theoretical predictions or results aggregate as well. In general, two considerations related to aggregation should be more prevalent in both TIEM and applied work:

1. What information is hidden when we aggregate individual treatment effects into an empirical target? In our example, the sign of the average treatment conveyed substantially less information about mechanisms than the individual treatment effects would provide (if they could be observed).

2. Can our choice of how to aggregate of individual treatment effects (i.e., the relevant operator) affect our observation of the causal process? In our example with mechanisms that generate counterveiling effects, quantile treatment effects (or potentially conditional average treatment effects [CATEs] for subgroups) provide more information about mechanisms than the average treatment effect.

When theories are articulated formally, mapping relevant contrasts into individual and then aggregate treatment effects facilitates an understanding of what can be learned from aggregate treatment effects. This exercise is rarely undertaken. But it is important because aggregation can obscure the mapping between a theoretical prediction and the empirical estimand.

5 Selecting the set of units for which causal effects are estimated

The remaining feature of the general causal estimand expressed in (1) is the set $\mathcal{D}$, which represents the set or profile of units for which treatment effects are estimated. Few—if any—theories in social science are universal in their scope. Whether our goal is use theory to interpret a measured treatment effect or use a treatment effect to test a theoretical prediction, we need to believe (some of) the units in $\mathcal{D}$ fall within the scope conditions of the proposed theory.

Understanding which units are or should be in the set $\mathcal{D}$ remains underspecified in applied work and in efforts to connect theory to credibility-oriented research designs. In principle, determining membership in $\mathcal{D}$ consists of three considerations that arguably cannot be collapsed further. First is a simple consideration of whether a theoretical mechanism or set of mechanisms can present in a given context or setting. A related consideration, sampling, deals with how researchers select units in order to detect or magnify the mechanism(s) of interest. Second, the effect of a mechanism on an outcome can also be magnified or diluted within specific subgrouops of units. Finally, some

credibility-oriented research designs limit researcher choices with respect to how to specify $\mathcal{D}$. To date, these considerations are most clearly articulated in work on external validity or generalization (e.g., Gailmard, 2021; Slough and Tyson, 2024). However, these considerations are also relevant for work in a single setting or on a single sample, as I discuss below.

5.1 Choosing settings and sampling units

When linking theory to empirics, researchers seek an alignment between the theoretical or conceptual environment and the empirical setting of a project. Typically, this is done by specifying the scope conditions of a mechanism (or set of mechanisms) within a theory. Ideally, the empirical setting falls within these conditions. If this were not the case, an important commensurability problem arises in which a treatment effect does not measure the influence of the posited mechanism(s). This limits the ability of our ability to test a theoretical prediction or use the theory to provide interpretation for the measured effect.

Scope conditions are often characterized as some contextual feature that apply to all subjects or units in a study (Slough and Tyson, 2023). But this need not be the case. More specifically, mechanisms may operate at different levels. A mechanism could be activated for individual subjects or all subjects within a certain context. It is not necessary for a given mechanism to be active for *all* units. This means that only a (non-zero) subset of units need to fall within the scope conditions for a given mechanism. Because treatment effects measure aggregate causal effects, consideration of how many units may fall within the scope conditions for a mechanism is an important, if underappreciated part of efforts to link theory to empirics.

For example, consider "power to the people"-style interventions (e.g, Björkman and Svensson, 2009; Raffler, Posner, and Parkerson, 2020) articles, local healthcare workers—by organizing citizens to put collective pressure on these bureaucrats. The specific interventions include a bundle of attributes and may activate multiple mechanisms by which these interventions improve service provision. Consider two of these mechanisms: citizen knowledge of healthcare services and the collective action capacity of a community. In principle, it should be possible to increase knowledge of individual citizens. For example, there may be citizens who are highly knowledgeable in the ab-

sence of intervention; increasing their (measured) knowledge may be infeasible. For less-informed citizens, however, information conveyed in community meetings may beget greater knowledge. In contrast, altering a community's latent level of collective action capacity applies to communities and all individuals therein.

The above example suggests that mechanisms can be activated for different subsets of units. It is undoubtedly a priority to make sure that all theoretical mechanisms could be operative for some subjects or units. However, in many cases, researchers will be unlikely to ensure that all units fall within the scope conditions for a given mechanism. In these cases, sampling decisions—within a given setting—can increase the possibility of detecting the effect of a mechanism. For example, if researchers measuring "power to the people"-style interventions wanted the best chance of observing the informational mechanism, they may want to (a) target communities with low aggregate levels of knowledge; or (b) (symmetrically) oversample citizens who are likely to have learned about healthcare services.[11]

The causal estimands championed by the credibility revolution provide a clear way to think through the mapping between scope conditions of a mechanism and researchers' choice of setting and units (e.g., experimental subjects) within that setting. We want to choose settings in which the mechanisms of interest can be observed for at least some individuals. Because we can only estimate aggregate treatment effects, however, we may want to use sampling as a tool to maximize our probability of observing the effect of a given mechanism when that mechanism is only likely to be active for a subset of units. Unfortunately, linking sampling considerations to theorized mechanisms is not standard practice. Because many scholars are worried about the generalizability of causal effects, any effort to improve our ability to observe manifestations of a mechanism could come at a cost (Egami and Hartman, 2020). Yet, if our goal is to activate a new mechanism to measure its influence, it is not clear that generalizability or, conversely, convenience, are the most persuasive rationales for a sampling strategy.

[11] Symmetry refers to using the same sampling strategy in treatment in control to maintain the internal validity of the experiment.

5.2 Specifying subgroups

When a given mechanism is thought to be activated (or its scope conditions satisfied) for only a portion of the sample, empiricists can also focus on estimating treatment effects for a subset of units. The resultant conditional (or subgroup) causal effects may provide a clearer measure of a mechanism's influence. When researchers specify subgroups, they choose the set $\mathcal{D}$ over which treatment effects are estimated.

For example, suppose that an experimental intervention is thought to activate a single mechanism at the level of individual subjects. However, some portion of the experimental sample falls outside the scope conditions for that mechanism. The ATE can then be construed as a weighted average of the effect of the mechanism[12] for those for whom it is active and zero effect for those for whom it is inactive. Suppose that we could predict the subset of units for whom the mechanism is thought to be active *ex-ante* on the basis of some pre-treament covariate or measured parameter of a model. Estimating the conditional ATE (CATE) for this subgroup should, in principle, come closer to isolating the influence of the mechanism when it is activated.

Estimating treatment effects for subgroups is particularly useful when directional theoretical predictions are mixed. Suppose that a given contrast is thought to produce positive effects on an outcome (via some mechanism or set of mechanisms) for some units but negative effects on that outcome (via a different mechanism or set of mechanism) for other units. The sign of the resultant ATE will depend on the magnitudes of these respective effects as well as the proportion of units in each subgroup. In this case, the ATE is not easily interpretable in terms of the mechanisms of interest since an ATE of any sign and magnitude could be consistent with the theorized mechanisms.

In applied work (Slough, 2024), I argue that bureaucratic quality affects the efficiency through which politician allocations to public goods yield improved public goods outcomes. This affects politicians' incentives to fund public goods, voter learning from observing public goods, and voter

[12]More precisely, it is the effect of the mechanism for a given contrast and outcome (Slough and Tyson, 2024).

Design	Interpretation of $\mathcal{D}$ without additional conditioning
Difference-in-differences	Treated units (when estimating any variant of treatment effects on the treated)
Instrumental variables	Complier units that are treated if and only if assigned to treatment (in the context of a binary instrument, binary treatment)
Regression discontinuity design	(Hypothetical) units at the discontinuity in the running variable

Table 1: Constraints on $\mathcal{D}$ in the most common specifications of three common observational designs for causal inference

choices. In particular, I show that (distinct) pooling equilibria emerge at low and high levels of bureaucratic quality in which voter observation of politician allocations should not affect their voting behavior. As a consequence, when empirical studies estimate treatment effects by pooling over municipalities with different levels of bureaucratic quality, they should attenuate estimates of the effects of voter learning about a politician's type (since voters should not learn in the pooling equilibria). In this setting, I show, average treatment effects attenuate the effects of voter learning. Instead, disaggregating treatment effects by level of bureaucratic quality provides evidence voter learning.

5.3 Research design as a constraint

Some observational credibility-oriented research designs constrain the set of causal effects that are identified or estimable. In these cases, we should think of a selected research design on a constraint on researchers' ability to choose $\mathcal{D}$. For example, regression discontinuity designs permit estimation of local average treatment effects (LATEs) at the discontinuity. In this case, the condition "at the discontinuity" specifies the set $\mathcal{D}$. Table 1 describes these constraints for the typical use of three observational designs. These constraints are methodological in nature and do not necessarily correspond to the theoretical considerations related to $\mathcal{D}$ that I have developed above.

When using these observational designs, researchers should consider the link between methodological constraints and theoretical considerations about how to best measure the effect of a given mechanism. In some settings, sample selection or partitioning units into subgroups may be an

effective strategy. In other settings, there may be a tension between the methodological constraints on $\mathcal{D}$ and our ability to use treatment effects to measure the effects of a mechanism. Additional research is needed to better develop these tensions.

6 Conclusion

I provide a new framework for organizing existing and future analyses of the commensurability of credibility-motivated empirical research and applied theory. The framework is organized around the causal effect estimands that credibility-motivated researchers seek to estimate. This framework shows how the credibility revolution provides important tools for furthering connections between theory and empirics in political science, the primary goal of the original EITM working group.

The framework reveals two areas ripe for further intervention in the EITM/TIEM tradition. First, because all empiricists estimate aggregate rather than individual treatment effects, more theoretical attention to issues of aggregation is sorely needed. Second, theoretical considerations of which units should contribute to aggregate causal effects can facilitate our understanding of both mechanisms underlying causal effects and increase interpretability of the resultant estimates.

References

Abramson, Scott F., Korhan Koçak, and Asya Magazinnik. 2022. "What Do We Learn about Voter Preferences from Conjoint Experiments?" *American Journal of Political Science* 66 (4): 1008–1020.

Angrist, Joshua D., and Jorn-Steffen Pischke. 2010. "The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics." *Journal of Economic Perspectives* 24 (2): 3–30.

Ashworth, Scott, Christopher R. Berry, and Ethan Bueno de Mesquita. 2015. "All Else Equal in Theory and Data (Big or Small)." *PS Political Science* 48 (1): 89–94.

Ashworth, Scott, Christopher R Berry, and Ethan Bueno de Mesquita. 2021. *Theory and Credibility: Integrating Theoretical and Empirical Social Science*. Princeton University Press.

Ashworth, Scott, Christopher R. Berry, and Ethan Bueno de Mesquita. 2024. "Modeling Theories of Women's Underrepresentation in Elections." *American Journal of Political Science* 68 (1): 289–303.

Björkman, Martina, and Jakob Svensson. 2009. "Power to the people: evidence from a randomized field experiment on community-based monitoring in Uganda." *The Quarterly Journal of Economics* 124 (2): 735–769.

Bueno de Mesquita, Ethan, and Scott A Tyson. 2020. "The commensurability problem: Conceptual difficulties in estimating the effect of behavior on behavior." *American Political Science Review* 114 (2): 375–391.

Clark, William Roberts, and Matt Golder. 2015. "Big Data, Causal Inference, and Formal Theory: Contradictory Trends in Political Science?" *PS Political Science* 48 (1): 65–70.

Egami, Naoki, and Erin Hartman. 2020. "Elements of external validity: Framework, design, and analysis." *Design, and Analysis (June 30, 2020)* .

Eggers, Andrew. 2017. "Quality-Based Explanations of Incumbency Effects." *Journal of Politics* 79 (4): 1315–1328.

Frangakis, Constantine E., and Donald B. Rubin. 2002. "Principal Stratification in Causal Inference." *Biometrics* 58 (1): 21–29.

Franzese, Robert. 2020. *The SAGE Handbook of Research Methods in Political Science and International Relations*. London: SAGE Publications chapter Econometric Modeling: From Measurement, Prediction, and Causal Inference to Causal-Response Estimation, pp. 577–598.

Fu, Jiawei, and Tara Slough. 2024. "Heterogeneous Treatment Effects and Causal Mechanisms." Working paper, New York University.

Gailmard, Sean. 2021. "Theory, History, and Political Economy." *Journal of Historical Political Economy* 1 (1): 69–104.

Hainmueller, Jens, and Daniel J Hopkins. 2014. "The Hidden American Immigration Consensus: A Conjoint Analysis of Attitudes toward Immigrants." *American Journal of Political Science* 00 (0): 1–20.

Hainmueller, Jens, Daniel J. Hopkins, and Teppei Yamamoto. 2014. "Causal Inference in Conjoint Analysis: Understanding Multidimensional Choices via Stated Preference Experiments." *Political Analysis* 22: 1–30.

Holland, Paul W. 1986. "Statistics and causal inference." *Journal of the American statistical Association* 81 (396): 945–960.

Huber, John D. 2017. *Exclusion by Elections: Inequality, Ethnic Identity, and Democracy*. New York: Cambridge University Press.

Izzo, Federica, Torun Dewan, and Stephane Wolton. 2020. "Cumulative knowledge in the social sciences: The case of improving voters' information." *Available at SSRN 3239047* .

Leamer, Edward E. 1983. "Let's take the con out of econometrics." *The American Economic Review* 73 (1): 31–43.

Lundberg, Ian, Rebecca Johnson, and Brandon M. Stewart. 2021. "What is your estimand? Defining the target quantity connects statistical evidence to theory." *American Sociological Review* 86 (3): 532–565.

Madestam, Andreas, Daniel Shoag, Stan Veuger, and David Yanagizawa-Drott. 2013. "Do Political Protests Matter? Evidence from the Tea Party Movement." *Quarterly Journal of Economics* 128 (4): 1633–1685.

Marshall, John. 2022. "Can Close Election Regression Discontinuity Designs Identify Effects of Winnign Politician Characteristics." *American Journal of Political Science* Early View: 1–17.

McConnell, Sheena, Elizabeth A. Stuart, and Barbara Devaney. 2008. "The Truncation-by-Death Problem: What to do in an Experimental Evaluation When the Outcome is Not Always Defined." *Evaluation Review* 32 (2): 157–186.

Montgomery, Jacob M., Brendan Nyhan, and Michelle Torres. 2018. "How Conditioning on Posttreatment Variables Can Ruin your Experiment and What to Do about It." *American Journal of Political Science* 62 (3): 760–775.

Neyman, Jerzy. 1923. "Sur les applications de la theorie des probabilites aux experiences agricoles: essai des principes (Masters Thesis); Justification of applications of the calculus of probabilities to the solutions of certain questions in agricultural experimentation. Excerpts English translation (Reprinted)." *Statistical Science* 5: 463–472.

Political Science Program, Directorate for Social, Behavioral and Economic Sciences. 2001. The Empirical Implications of Theoretical Models (EITM) Workshop. Report National Science Foundation.

Raffler, Pia, Daniel N. Posner, and Doug Parkerson. 2020. "Can Citizen Pressure be Induced to Improve Public Service Provision?" Working paper, Harvard University.

Rubin, Donald B. 1974. "Estimating causal effects of treatments in randomized and nonrandomized studies." *Journal of educational Psychology* 66 (5): 688.

Samii, Cyrus. 2016. "Causal empiricism in quantitative research." *The Journal of Politics* 78 (3): 941–955.

Slough, Tara. 2023. "Phantom Counterfactuals." *American Journal of Political Science* 67 (1): 137–153.

Slough, Tara. 2024. "Bureaucratic Quality and Electoral Accountability." *American Political Science Review* FirstView: 1–20.

Slough, Tara, and Scott A. Tyson. 2023. "External Validity and Meta-Analysis." *American Journal of Political Science* 67 (3): 440–455.

Slough, Tara, and Scott A. Tyson. 2024. *External Validity and Evidence Accumulation*. New York: Cambridge University Press.

Wolton, Stephane. 2019. "Are Biased Media Bad for Democracy?" *American Journal of Political Science* 63 (3): 548–562.

Wolton, Stephane. 2024. "TIEM.".
URL: *https://stephanewolton.com/about/tiem/*