Electoral Studies 40 (2015) 372–393

Contents lists available at ScienceDirect

Electoral Studies

journal homepage: www.elsevier.com/locate/electstud

EITM: An assessment with an application to economic voting☆

Jim Granato [a,*], Cong Huang [a], Kwok Wai Wan [a], Ching-Hsing Wang [b], Man Chiu Sunny Wong [c]

[a] Hobby Center for Public Policy and Department of Political Science, University of Houston, 306 McElhinney Hall, Houston, TX 77204-5021, USA
[b] Hobby Center for Public Policy, University of Houston, 306 McElhinney Hall, Houston, TX 77204-5021, USA
[c] Department of Economics, University of San Francisco, 2130 Fulton Street, San Francisco, CA 94117-1080, USA

ARTICLE INFO

Article history:
Received 9 May 2014
Received in revised form 24 February 2015
Accepted 2 March 2015
Available online 20 March 2015

Keywords:
Empirical implications of theoretical models
Economic voting
Formal modeling
Empirical modeling
Error-in-variables regression
Conditional expectations

ABSTRACT

This paper describes the origins of EITM and its role in research and training. The EITM framework is reviewed as well as clarifications of its meaning. We apply EITM to the study of economic voting and demonstrate how linking formal and empirical analysis creates a dialogue between theory and test sufficient in supporting communication and cumulation of scientific knowledge.

1. Introduction and background

On July 9th and 10th, 2001 the Political Science Program at the National Science Foundation (NSF) convened a workshop seeking ways to improve technical-analytical proficiency in Political Science.[1] This workshop, termed the Empirical Implications of Theoretical Models (hereafter EITM) Workshop, suggested constructive approaches so the NSF Political Science Program could develop linkages and a dialogue – both methodological and interpersonal – between formal and empirical modeling.

The participants in the workshop – with diverse methodological backgrounds – were senior scholars with research experience in various technical-analytical areas and proven track records in activities that have improved the technical-analytical expertise in various social sciences. Participants were primarily from political science, but economics and mathematics were represented as well.[2]

☆ We thank Frank Scioli for his assistance. This paper containing complete references can be found at: http://www.class.uh.edu/hcpp/EITM/institute.htm.

* Corresponding author.
E-mail address: jgranato@central.uh.edu (J. Granato).

[1] The 2001 NSF EITM Workshop was recorded and transcribed. The written transcript is available on the NSF Political Science Program Web Site: www.nsf.gov/sbe/ses/polisci/reports/eitm709.pdf and www.nsf.gov/sbe/ses/polisci/reports/eitm710.pdf. A report of the EITM initiative – based in part on the 2001 EITM Workshop – (EITM Report 2002) is also available at: www.nsf.gov/sbe/ses/polisci/reports/pdf/eitmreport.pdf.

[2] The participants and their commentaries are listed in Appendix B of the 2002 EITM Report.

http://dx.doi.org/10.1016/j.electstud.2015.03.003
0261-3794/

1.1. Why EITM was created: motivation, problem diagnosis, and remedies

There were several motivating factors that led to the EITM initiative at NSF, ranging from reputational to matters of basic research design and competitiveness of grant submissions. EITM was believed to be one way to reverse or at least combat the following issues:

- Change the negative image of the political science discipline at NSF
- Address old methodological antagonisms that negatively influence methodology training
- Address fundamental reasons that lead to noncompetitive NSF proposals.

Motivation 1. Perceived Weakness of the Political Science Discipline at NSF.

Granato and Scioli (2004) cite the following report relating how political science was perceived at NSF:

> The recent *Report of the APSA Ad Hoc Committee on the National Science Foundation* found that political science had been characterized as, "not very exciting, not on the cutting edge of the research enterprise, and in certain quarters as journalistic and reformist."[3] We disagree with this statement and believe there has been considerable improvement in political science in the past 40 years through the use of formal models, case studies, and applied statistical modeling (page 313).

This negative perception also led to skepticism as to whether the political science discipline (i.e., its current training practices) was technically equipped to improve upon the existing methodological status quo. Social and Economic Sciences (SES) Division Director Bill Butz stated all was not certain about the outcome:

> Sometimes that works and sometimes you're just pushing on a string because the field isn't ready for it yet … And getting you all here and I judge from the papers it resonated with you, too. And we'll see in the succeeding year or 2 or 3 whether this is pushing on a string or whether it's really lighting a fire (EITM Workshop Transcript 2001: 18).

Motivation 2. Old Antagonisms and the Methodological Status Quo.

Workshop participants were from varied methodological backgrounds where long antagonisms had existed and led to splits in departments as well as various subfields. But, EITM workshop panelist Dina Zinnes expressed hope that these old antagonisms between formal and empirical modelers could be overcome and lead to some meaningful advice.

> First let me just say what a pleasure it is to be amongst this group of people. I have to admit that when I got those initial memos I sort of put them on the side burners, thinking, well, okay, I'll look at them eventually, because I was worried about the fights and the antagonisms that I thought would emerge. And it was with great delight that I read those and discovered, my gosh, there really is a consensus going on here. And listening to people this morning confirms that. I find that it's wonderful to see that both the empirical and statistical side and the modeling side really all sort of agree on certain things. And I think that's a fabulous beginning (EITM Workshop Transcript 2001: 113–114).

Motivation 3. Weaknesses in Research Design for NSF Competitions.

In his role as SES Division Director over a six year period, Director Butz reviewed and approved over 16,000 proposals. He stated:

> And of those 16,000, about 2 years ago I formulated just a sort of a stylized FAQ what the principal ways are to be sure that you don't get money from NSF. And out of all the possible reasons, there were three that came to the front … Now, it varies some across fields. And I don't mean to say that this is particularly true of political science, but I want to show it to you because it may give you an additional context for the reasons why scientific proposals fail in the social and behavioral sciences – how to get zero money (EITM Workshop Transcript 2001: 14).

One reason for noncompetitive proposals – in Director Butz's survey – is that they are vague in conceptualization. There is "no sense of how what this person is doing fits into what came before conceptually or how the results, if they're confirmed or not confirmed, will feed some kind of a general conceptual sense of what's going on (EITM Workshop Transcript 2001: 14)."

A second reason is even though basic conceptualization exists, there is still a failure to connect theories to tests:

> there will be a well-developed deductive theory at the beginning, and then the next section will be data, the next section will be empirical equations, and you'll look at the empirical stuff and it's just – it's not connected, or it's only connected in the vaguest sense (EITM Workshop Transcript 2001: 14–15).

A final reason in his summary was inadequate specification:

> I don't know how many panels I've sat in where people say, well, you know, we can't really tell how they're going to form this proxy from these variables, or we can't really tell how they're going to get over the statistical problem with such-and-such (EITM Workshop Transcript 2001: 17).

In concluding his presentation Director Butz states:

> There are many other things that are wrong with proposals, but these two – something wrong with the theory and something wrong with the data or the statistical methods are two of the three most common ones

[3] Report of the APSA Ad Hoc Committee 2000, page 1. American Political Science Association 1527 New Hampshire Avenue, NW Washington, D. C. 20036.

> across – and I really don't think there are very many exceptions to this – across the 18, I think now 19, programs in the social, behavioral, and economic sciences here. So I thought I would just point that out (EITM Workshop Transcript 2001: 16–17).

1.1.1. Problem diagnosis: siloed training and thinking in methodology

Against this backdrop it was clear social science methodology and the attendant research designs that are based on said methodology played an important role in the negative outcomes above, not only for political science, but for other social and behavioral sciences. While research methodology encompasses many elements, the EITM workshop focused on the state of quantitative methodology and whether it was a source of the problems cited above.[4]

In diagnosing the factors workshop participants reflected on the natural division of labor and specialization, but also the cost of this natural division:

> Isolation – ***compartmentalization*** – of fields and subfields is the *status quo* in political science … This current field and sub-field structure exacerbates the separation between formal and empirical modeling. For example, focusing on a question that is particular to American Politics increases specialization and, turn, discourages integrating approaches and theories that would best come about from studying a particular research question in many countries (EITM Report 2002: 6).

Moreover, field and sub-field isolation reinforces separation between formal and empirical analysis including the belief that an:

> outdated perspective about formal and empirical analysis is the assertion that these technical-analytical approaches are simply interesting intellectual enterprises that lack political and social relevance (EITM Report 2002: 6).

The consequence of this divide is not neutral in its effect; indeed the effect can be negative. In particular:

> a good deal of research in political science is competent in one technical area, but lacking in another, that is, a formal approach with substandard (or no) empirical tests or an empirical approach without formal clarity. Such impaired competency contributes to a failure to identify the proximate causes explicated in a theory and, in turn, increases the difficulty of achieving a meaningful increase in scientific knowledge (EITM Report 2002: 1).

To sum up, a "siloed" research program contributes to a failure to identify the proximate causes explicated in a theory and, in turn, increases the difficulty to achieve a meaningful increase in scientific knowledge. Researchers limit themselves to the strengths and weaknesses of a single methodological approach. For formal modelers, this manifests itself in not respecting the facts; for researchers who rely exclusively on applied statistics, we find data mining, garbage-can regressions, and statistical patches (i.e., omega matrices) (Granato and Scioli, 2004).

1.1.2. Siloed training: consequences for formal and empirical modeling

Siloed training in formal modeling can be sourced to basic comfort levels in approach:

> Many formal modelers feel uncomfortable with powerful empirical concepts such as social norms, limited rationality, and psychological factors such as personality and identity.[5] The usual argument is that formal models are not meant to fit data, or should not be. While there is much to be learned from pure theory and abstract formal arguments, the formal modeling isolation reinforces distance from basic circumstances that these abstract models could help to illuminate. This isolation also contributes to the basic misunderstanding noted above about the great attributes formal modeling brings to the scientific process (EITM Report 2002: 6–7).[6]

Empirical modelers face their own limitations:

> Empirical modeling isolation, on the other hand, is equally guilty of not advancing scientific understanding when it fails to incorporate their "more complex and general assumptions" into a mathematically identified model with direct and testable implications. Instead "errors" or "confounding variables" that derail the inferential process are treated as statistical problems that require only statistical fixes (EITM Report 2002: 7).

1.1.3. Factors reinforcing the status quo

The resistance to unify formal and empirical modeling is due to several factors. These obstacles are not only contained in the 2002 EITM Report but more recently stated in Poteete et al. (2010: 3–27). Among those factors:

The Intellectual Investment: Scholars have to invest in different skill sets. The intellectual investment needed for formal modeling is different than the knowledge needed for empirical modeling. But, given the greater mathematical demands in formal modeling the tendency is for students and scholars not to have sufficient training in formal

[4] The EITM initiative is part of a multi-method approach. Recall the motivation for the use of EITM has quantitative roots, but consistent with the arguments of Poteete et al. (2010), it is recognized that qualitative approaches have various strengths, including highlighting the importance of context (Granato and Scioli (2004: 314–315). But, as with quantitative tools, Granato and Scoili (2004) do highlight shortcomings of qualitative approaches.

[5] From the 2002 EITM Report:

> A good example of the consequences of formal modeling isolation can be found in psychology. Despite a growing literature in mathematical psychology, a perusal of the Journal of Mathematical Psychology reveals that mathematical modeling tends to be limited to the simplest of individual learning and perceptual phenomena (page 6).

[6] The argument here should not be confused with pure theoretical work that informs subsequent theoretical work. Arrow's theorem (1963) is a case in point. His model is based on well-established logical properties. Data are not necessary here. In particular, he uses the properties for equivalent relations in preference functions (i.e., reflexivity, symmetry, transitivity) and sets up five minimum conditions – which seem quite reasonable – for social choice aggregation. He finds these conditions cannot be satisfied when aggregating individual orderings. This work has been extended to include refinements to the theory – cumulation (Clark and Primo, 2012: 85).

modeling. This deficit is compounded since there are few incentives to motivate tenured faculty to try new methods, including using formal modeling as part of their tool kit.

Training Differences: Empirical modelers devote their energies to data collection, measurement, and statistical matters, and formal modelers focus on mathematical rigor. This divide is reinforced in departments having a strong tradition in either formal or empirical analysis but not both.

Research Practice: For empirical modelers, model failures lead to emphasis on additional statistical training or more sophisticated uses of statistics – usually to "patch over" – a model failure. Formal modelers, on the other hand, deal with model controversies by considering alternative mathematical formulations but this is usually done piecemeal. A similarity between these two approaches is both formal and empirical modelers tend to remain tied to their particular technique despite the warning signals evidenced in model breakdown. These practices are reinforced by reviewers and journal editors because of their specialization in either formal or empirical analysis.

These implementation challenges are deeply rooted in the academic community – fostered by career incentives – taking years to overcome (Poteete et al., 2010: 18–24). Consequently, "old habits" learned in graduate school inhibit the desire to make the changes in skill development. But, the situation is worse since many things learned in graduate school tend to become out-of-date by mid-career.

When methodological instruction reflects these status quo forces, successive generations will only repeat the shortcomings. Indeed, disciplines failing to provide incentives for this type of risk taking and re-tooling increase the occurrence of an:

> assembly-line model of research production that imperils innovative theories and methodologies and, in turn, scientific breakthroughs. One could make the argument that EITM or initiatives like it are unnecessary because the unfettered marketplace of ideas expedites best scientific practices and progress. But, it is precisely because there are significant rigidities (training and otherwise) in the current academic setting (imperfect competition) which makes EITM-type initiatives not only necessary – but imperative (EITM Report 2002: 8).

We now see, and have repeatedly seen, practices unsuitable for addressing complex issues. The failure to build cumulative research areas is not the only consequence either. Invalid policy prescriptions take place: prediction without basic understanding of how a system under study works is of little scientific or social value.

1.1.4. Proposed remedies

In both written and spoken commentaries, EITM Workshop participants recommended the NSF Political Science Program address the technical-analytical divide between formal and empirical approaches in three priority areas:

- Education: Training and Retraining
- Dissemination of Knowledge: Conferences and Workshops
- Research: Establishment of Research Work Groups.

The EITM initiative, then, was viewed as using these priority areas to expose students to integrating formal and empirical approaches. Students (and faculty) gain a vantage point and a means to escape the destructive practices resulting from siloed training and research. By integrating the two approaches students would be exposed to the strengths of both approaches:

> At the most basic level, formal modeling assists in the "construction of valid arguments such that the fact or facts to be explained can be derived from the premises that constitute the explanation."[7] An important virtue of formal modeling is that it often yields surprising implications that would not have been considered had they not emerged from formal analysis. Conversely, if practiced correctly, applied statistical and case study analysis shows researchers where a model went wrong and leaves open the possibility that a more accurate model can be constructed (Granato and Scioli, 2004: 314).

1.2. Deliverables from the 2001 EITM workshop

The proposed remedies were circulated – in the form of a "Dear Colleague Letter" – approximately 3 weeks after the Workshop concluded and covered a call for establishing EITM summer training institutes, workshops, and assembling research work teams. The call was answered and the first competition was completed in March, 2002 with the first EITM activities underway in the summer of 2002. There have been subsequent competitions for the summer training institutes and a one-time only EITM graduate fellowship program that was competed in fiscal year 2003.

A key outcome of the EITM initiative has been the EITM Summer Institutes. The Summer Institutes have taken place at:

- Harvard University (2002).
- The University of Michigan (2003, 2006, 2009).
- Duke University (2004, 2008, 2014).
- UC-Berkeley (2005, 2010, 2013).
- UCLA (2007).
- Washington University, St. Louis (2003–2009).
- University of Chicago (2011).
- University of Princeton (2012).
- University of Houston (2012–2014).

Generally, the EITM Summer Institutes cover two to four weeks of intensive training (the two- and three-week institutes often provide training six-days per week) with morning and afternoon instructional presentations, and evening laboratories or workshops where participants complete their daily assignments.

Since the inception of the EITM initiative, approximately 450 students graduated from the Summer Institutes, both stipend and non-stipend. If the data are broken down further the reach of the program may be understated. For example, Washington University reported that, for the 2003–2009 period, there were 163 stipend participants,

[7] See Wagner (2001: 3).

and at least 75 non-stipend. The latter included about a dozen from European universities as well as a large number of Washington University Ph.D. students, some of whom only participated in certain sessions but some of whom completed the whole program.

1.3. Evaluation: the 2009 EITM workshop

In 2009, the NSF Political Science Program convened a second workshop asking faculty participants to evaluate the impact of the EITM initiative and, more specifically, the summer institutes.[8] From their oral and written commentaries, the 2009 Workshop participants indicated that EITM had a positive scientific impact during the past decade. They noted the support and participation of many prominent scholars in various components of the EITM initiative, including such outstanding social scientists as 2009 Nobel Laureate Elinor Ostrom.

The 2009 Workshop also assessed the impact of the summer institutes. The data from an e-mail survey conducted by Washington University of past student participants in its institutes showed a positive effect of the institute on the participants' future progress. For example, 36 out of 43 respondents indicated that the institute played an important role in framing their dissertation projects, and 11 engaged in further collaboration with other EITM participants. Furthermore, 23 of the 43 EITM graduates who participated in the survey went into tenure-track faculty positions.

Similarly, an e-mail survey of participants of the first rotating summer institutes (Harvard, Duke, Michigan, UCLA, UC-Berkeley) found (*at the time of the survey*) that 83 currently hold tenure-track assistant professor positions, five hold tenured associate or full professor positions, six were currently completing post-doctoral fellowships, three have other research positions, and nine are still students (the remaining 33 did not respond to the survey).[9]

1.4. EITM graduates

One purpose of the EITM initiative is to strengthen research abilities and competitiveness of junior scholars and graduate students. As mentioned by Aldrich et al. (2008: 829–830):

> This [EITM] training not only educates students about highly-technical theoretical and empirical work but encourages them to develop research designs that provide more precise and dynamic answers to critical scientific questions by integrating the two approaches from the very beginning of the work. Most importantly, perhaps, there is a critical mass of scholars (particularly in the discipline's junior ranks), who recognize how this approach can improve the reliability and credibility of their work.

The EITM training in over the past decade has been conducive to enhancing research ability of participating students. While it would be risky to ascribe student success primarily to their exposure to the EITM initiative, we do find many EITM graduates earning a tenure track faculty position after completing a doctorate degree.[10] Moreover, some EITM graduates are able to apply the EITM approach to their research. Among the research published in some prominent academic journals are:

- *American Political Science Review* (Ahmed, 2012)
- *American Journal of Political Science* (Bonica, 2013; Svolik, 2009)
- *American Journal of Sociology* (Centola et al., 2005)
- *The Journal of Politics* (Penn, 2008)
- *Journal of Peace Research* (Bapat, 2005), and
- *Journal of Theoretical Politics* (Arena, 2015).

The remainder of this paper is organized along the following lines. The following section provides the EITM framework developed at NSF. Sections three and four provide clarifications on the meaning of EITM as well as criticisms. The fifth section describes the relation between EITM and formal and empirical modeling dialogues. Section 6 discusses EITM limitations. An EITM application to economic voting is provided in section seven and the eighth section concludes.

2. The EITM framework

EITM is a method – even a mindset – where researchers treat formal and empirical analysis as linked entities intended to create a dialogue between theory and test. We have already demonstrated the motivation for EITM as a means to counter destructive training and research practices due to compartmentalization. We agree specialization and the division of labor are necessary to begin training processes, but it is also clear that integration is an important next step in the training process and eventually for research practice.

But, these motivations are not enough if we want to implement EITM. Implementation involves defining the elements of EITM – a framework – and showing how one does EITM research and how one trains students to do such research. The development of a framework is important since "without a framework to organize relevant variables identified in theories and empirical research, isolated knowledge acquired from studies … by … social and behavioral scientists is not likely to cumulate (Ostrom, 2009: 420)."

The EITM framework outlines how formal and empirical methods can be unified and how it assists in building a cumulative political and social science. We hasten to add this type of ***methodological unification*** is not new in the

[8] A report of the 2009 Workshop can be found at: http://www.class.uh.edu/hcpp/EITM/Literature/eitmreport_2010.pdf.

[9] An EITM certification program has also been created at the *University of Michigan's ICPSR Summer Program in Quantitative Methods of Social Research* (See http://www.icpsr.umich.edu/icpsrweb/sumprog/and http://www.eitminstitute.org/index.html). The certification program requires students focus on a set of "approved" courses that provide background for using the EITM approach and attending the EITM Summer Institutes. Between 2011 and 2013 over 270 students have received certification (See http://www.eitminstitute.org/recipients.html).

[10] More detailed information about the EITM graduates' current institution can be found at: http://www.eitminstitute.org/alumni.html.

social sciences. It can be traced back to the accomplishments of the Cowles Commission.[11] The Cowles Commission approach put emphasis on structural equation modeling as well as conditions for parameter identification.[12]

The EITM framework builds on the Cowles Commission approach and then places an emphasis on *developing behavioral and applied statistical analogues and linking these analogues*.[13] The framework includes the following attributes:

1. EITM places emphasis on modeling human behavior so new uncertainty created by shifts in behavioral traits such as public tastes, attitudes, expectations, and learning are properly accounted for and studied.
2. The Cowles Commission is associated with building a system of equations and then following rules (rank and order conditions) for identification that count equations and unknowns. Our EITM framework is agnostic on the choice to build and relate a system or to partition the system (via assumption) into a smaller set of equations, even a single equation. This debate about general and partial equilibrium model building can be traced back to at least the 1800s.
3. A final and related point on model specification relates to the critiques of the structural approach leveled by Sims (1980). It is well known that structural parameters are not identified from reduced form estimates. The practice of finding ways to identify models can lead to "incredible" theoretical specifications (Sims, 1980). The proposed EITM framework, by adding behavioral concepts and analogues, can address Sims' criticisms in a theoretically meaningful way. Analogues, in particular, have important scientific importance since they hold the promise of operationalizing mechanisms.[14]

This EITM framework contains three steps[15]:

Step 1. Relate and Unify Theoretical Concepts and Applied Statistical Concepts

The goal of this first step in EITM is to transform the focus from the substantive topic to the underlying behavioral process. We start, however, not with the development of mathematical structures but with the identification of concepts. It is of course standard to suggest that research start with concepts. This is not new in and of itself. We have in mind, however, not the substantive concepts central to a discipline, but instead to the general behavioral attributes of the thing being researched.[16]

Concepts of particular concern in this framework reflect many overarching social and behavioral processes. Examples include (but are not limited to):

- decision making
- bargaining
- expectations
- learning
- elements of social interaction (strategic and non-strategic)

It is also important to find an appropriate statistical concept to match with the theoretical concept. Examples of applied statistical concepts include (but are not limited to):

- persistence
- measurement error
- nominal choice
- simultaneity
- prediction (ex-ante, ex-post)

Step 2. Develop Behavioral (Formal) and Applied Statistical Analogues

Concepts and test linkage require analogues. Recall that an analogue is a device representing a concept via a continuous and measurable variable or set of variables. Examples of analogues for the behavioral (formal) concepts such as decision making, expectations, learning, and strategic interaction include (but are not limited to):

[11] Created in the 1930s, the Cowles Commission was designed "to foster the development and application of rigorous logical, mathematical, and statistical methods of analysis" for application in economics and related social sciences (See http://cowles.econ.yale.edu/about/index.htm).

[12] However, at that time there was a lack of emphasis on agent behavior, including responses to alternative policies, as well as other social, political, and economic factors. As a consequence, using original Cowles Commission practices, we cannot predict how the behavioral response of agents influence the success or failure of a policy or treatment. The reason, as Lucas (1976) has argued, is that in-sample estimation provides little guidance in predicting the effects of policy changes because the parameters of the applied statistical models are unlikely to remain stable under alternative stimuli.

[13] Analogues are related to concept operationalization. An analogue is a device represented by variable – and measurable – quantities. Analogues include variables, operators, or an estimation process that mimic the concept of interest. They serve as analytical devices – not categorical indicators – for behavior and, therefore, provide for changes in behavior as well as a more transparent interpretation of the formal and applied statistical model.

[14] An early example of operationalizing a mechanism can be seen in the work of Converse (1969). He advanced the theory that strength of party identification (and voting behavior) is a function of intergenerational transmission plus the number of times one had voted in free elections. To operationalize his proposed mechanism – intergenerational transmission – he made use of the following analogue: the Markov chain. This particular analogue allowed for a particular dynamic prediction he linked with data.

[15] See Granato (2005) and Granato et al. (2010a, 2010b, 2011) for a description and examples of the EITM framework.

[16] In political science, for example, a student of democracy might focus on choice (and decision making): how do demographic and attitudinal variables drive individual selection over political parties. Another student might focus on uncertainty and learning: given the lack of a "track record" among political parties in newly democratizing states, how do individuals come to form expectations regarding those parties, and how do those expectations shift in response to political and economic changes? A third student might concentrate on the idea of bargaining: how do the various party leaders face the trade-offs between maximizing their potential influence in the political system and maintaining the promise to democratize? The idea is not to ignore the substantive aspects, but to look at substance from a different perspective, one that not only helps clarify the focus of the research but also suggests common behavioral concerns that make it easier to communicate across subfields and find common elements and approaches. We thank Douglas Dion for this set of examples.

- decision theory (e.g., utility maximization)
- conditional expectations (forecasting) procedures
- adaptive and Bayesian learning (information updating) procedures
- game theory

Examples of applied statistical analogues for the applied statistical concepts of persistence, measurement error, nominal choice, simultaneity, and prediction include (respectively):

- autoregressive estimation
- error-in-variables regression
- discrete choice modeling
- multi-stage estimation (e.g., two-stage least squares) and spatial econometrics
- point estimates and distributions[17]

Step 3. Unify and Evaluate the Analogues

The third step unifies the mutually reinforcing properties of the formal and empirical analogues. By starting with the concept and then moving to the theoretical and applied statistical analogues, we guarantee that there must be something in common between the theory and the empirical analysis. The idea, then, is to locate the parameters of interest in each that reflect the underlying concept, and then use those to build clearer and stronger links between the mechanisms of the theoretical model and the specification of the statistical methods. The specified linkage not only draws theory and empirics closer, but also provides a way for research to build by showing potential sources of inaccuracies and model failure.

2.1. Summary

What the EITM framework offers is a form of methodological unification that: a) emphasizes basic behavioral concepts, and b) uses mutually reinforcing properties of formal and empirical analysis. The 2001 Workshop highlighted the key contribution of the EITM approach to both political science and the social sciences:

> If one were to summarize in one word what bridging the divide between formal and empirical modeling means for the political, social and behavioral sciences, that word would be ***identification***. The ability of a researcher to ***identify*** or parse out specific causal linkages among the many factors is fundamental to the scientific enterprise. Specifying a model that links both formal and empirical approaches alerts researchers to outcomes ***when specific conditions are in place*** – and is also one of the best ways to determine an ***identified*** relationship (pages 1–2).

This EITM framework should not be interpreted as a substitute for pure formal or pure empirical approaches. While the shortcomings of these approaches have been noted above, their strengths are evident. Their valid use is essential particularly when theory or data are either underdeveloped, nonexistent, or both. The simple fact is there are numerous examples in many sciences where theory is ahead of data or data are ahead of theory, sometimes for decades.[18] Nor should the quantitative nature of this framework suggest it precludes the use of qualitative procedures. Such exclusion would be throwing out information which could otherwise aid in finding underlying mechanisms.

3. EITM: clarifications

Two clarifications of the EITM initiative are in order. The first relates to whether EITM is a research initiative. The second clarification explains how using the EITM framework outlined above does not constitute a rejection of a division of labor in current training and research practices.

3.1. EITM is broader than a research initiative

In their book, *A Model Discipline*, Clarke and Primo (2012) assert that:

> The EITM research initiative is a National Science Foundation (NSF) funded project to develop formal models that are tested with data (page 48).

They also argue:

> The goal of the initiative is to “bridge the gap” between formal and empirical analysis (EITM Report 2002: 5) (page 48).

These statements by Clarke and Primo do not fully reflect the core idea of EITM. EITM is not just a research initiative. It is an initiative that affects both training (education) and research.[19] Moreover, as stated in the 2002 EITM Report, “EITM opportunities for education (training), knowledge transmission, and research work teams are designed to bridge the gap between formal and empirical analysis by addressing the factors that have produced that gap (page 5).”

Although Clarke and Primo correctly point out that one of the ideas of EITM is to bridge the gap between formal and empirical analysis, we need to stress this is the *consequence*, not the *cause*. The “gap” can be closed when researchers are willing to undertake actions to overcome the shortcomings of the research practices. Readers can also refer to Section 1 above, but further clarification can be found in the following passage:

> What can be done? One way to address these problems is to change standard research practices and enhance training opportunities so that formal and empirical analyses (applied statistical *and* case study analyses) are

[17] While we focus here on applied statistical analogues, there are other types of tests, including simple numbers and qualitative outcomes in case studies and experimental settings. One could think of these other analogues as a bridge to successive tests where better data and tools become available.

[18] The EITM framework allows for multiple ways to test a formal model. This includes (but is not limited to) experiments, secondary data, primary data, case studies, a formalization showing a link to a test when applicable data are non-existent, or a combination of these and other tests.

[19] Note also a particular interest arising from the 2009 NSF EITM Workshop is the impact on training and research, and specifically how EITM *leads to a reorientation of training and research*.

linked. Large-N analysis can test a formal model through statistical analysis, and small-n case studies can also test a model by seeing if the mechanism postulated by the model really exists … (Granato and Scioli, 2004: 314).

To be even more specific, the programmatic initiatives implementing the EITM initiative are found in the 2002 EITM Report (page 10):

> To address the skills deficit in formal modeling, empirical modeling, and especially both, support can be provided for graduate training, post-doctoral opportunities, and mid-career re-tooling. Such support can include, but is not limited to, courses in formal and empirical modeling. For graduate students, funding could be provided for an additional year or two of graduate school to complete both formal and empirical modeling sequences. For faculty, support could be given to visit another department on campus or another institution.
>
> Support can also consist of summer training institutes and training centers that are positioned to serve larger numbers of individuals while reaching graduate students and faculty who are in departments that cannot offer this training. These individuals become exposed to more experienced social and behavioral scientists who combine formal and empirical analysis. The forms of exposure can vary, ranging from a summer (semester) to shorter-term lectures or workshops (one-week).

3.2. EITM and the division of labor in training and research

The negative effect of specialization and compartmentalization is discussed in Section 1. However, it would be a mistake to say the EITM initiative does not comprehend that a division of labor, particularly in early stage training courses, is important. Recall that each method has its own strengths and weaknesses. It is not the division of labor that is the problem; rather, it is the methodological isolation which is harmful. EITM is meant to break the status quo in siloed training approaches. This is wholly different than respecting a division of labor for foundational training courses.

It is important to reiterate the strengths each approach possesses.[20] Formal models, for example:

> … force clarity about assumptions and concepts; they ensure logical consistency, and they describe the underlying mechanisms that lead to outcomes. They also can lead to surprising results, such as the free rider problem or the power of the median voter, which have spawned substantial literatures (Granato and Scioli, 2004: 313).

For empirical models, they:

> … can provide generalizations and rule out alternative explanations through multivariate analysis. Researchers are forced to conceptualize putative causes so that they can be reliably measured. Models can distinguish between causes and effects, allow for reciprocal causation, and estimate the relative size of effects (Granato and Scioli, 2004: 314).

However, the division of labor must be understood in the context that each method has shortcomings which can be reinforced:

> What we find is that because they are generally treated by scholars as distinct, separable approaches, the three most common current research practices – formal modeling, case study analysis, and applied statistical modeling – deviate from this ideal. They therefore limit the possibilities for substantial enhancement of knowledge (Granato and Scioli, 2004: 315).

In particular,

> formal models can fail to incorporate empirical findings in order to provide a more accurate depiction of the specified relations. The models may be elegant, but too often they ignore, or even throw out, useful information. This results in modeling efforts that yield inaccurate predictions or do not fit findings. In fact, data may contradict not just a model's results but also its foundational assumptions (Granato and Scioli, 2004: 313).

Empirical analysis, when it degenerates into being the only tool employed, has left some:

> … of the best methodologists to wonder if we have reached the point of diminishing marginal returns with statistical analysis. The variables in regressions are sometimes poor reflections of theoretical concepts. Empirical models often seem more like data mined "garbage-can regressions and garbage-can likelihoods" because of their lack of causal motivation and theoretical specificity.[21] Indeed, model shortcomings are typically treated as statistical problems requiring statistical patches, instead of a more careful specification of the mechanism behind the model. The distance between theory and test can only grow with this mindset (Granato and Scioli, 2004: 314).

[20] As mentioned earlier, Granato and Scioli (2004) also discuss the role of qualitative analysis in a research design. While EITM is about the unification of formal and empirical analysis, qualitative analysis also has particular strengths which aid both theory and empirics. They include:

> … detailed information about the steps by which events occur and allow researchers to identify mechanisms that can produce such phenomena as group-think, authoritarian regimes, revolutions, and ethnic conflict. Case studies also enable researchers to discover enough about countries to distinguish idiosyncratic from general causes, to identify interactive and connected causes, and to understand how people's interpretations of events – the meaning that they have for people – affect their actions (Granato and Scioli, 2004: 314).

On the other hand, qualitative analysis shortcomings occur when the focus is:

> … too much on the idiosyncratic details of rare and influential events. They may miss the opportunity to inform a more general theory. In some instances the result amounts to theorizing by proverb: that is, site-specific theories expressed as causal theories (Granato and Scioli, 2004: 314).

[21] Achen (2002: 424) states these empirical models (or practices) "are too often long lists of independent variables from social psychology, sociology, or just casual empiricism, tossed helter-skelter into canned linear regression packages." He credits Anne Sartori for the term "garbage-can regressions."

4. EITM criticisms

Various criticisms have been leveled at the EITM initiative. The criticisms center on: the motivation for EITM and the degree to which EITM is related to the hypothetico-deductive (H-D) approach. *At this point in time we find no criticisms leveled at the EITM framework outlined in Section 2.*[22] *Rather, criticisms have been leveled at statements in* Granato and Scioli (2004) *and the documents related to the 2001 NSF EITM Workshop (EITM Report 2002).*

4.1. The motivation for EITM

Motivation criticisms of EITM have appeared in Clarke and Primo (2012). We discuss three motivations in Section 1, but here we present Clarke and Primo's arguments and their quotes of various EITM sources. We respond using the same sources.

Clarke and Primo's criticisms on the motivation for EITM can be found in this quote:

> For all the discussion of science in Granato and Scioli's article, there is no actual argument or justification for pursuing an EITM strategy. We are told, for instance, that "by thinking about the empirical implications of theoretical models scholars develop clear-cut empirical tests (Granato and Scioli, 2004: 314). Although the statement may or may not be true, it is not an argument that scientific progress results from developing clear-cut empirical tests. That conclusion is simply assumed (Clarke and Primo, 2012: 50).

This statement is false. There are many motivations for "pursuing an EITM strategy" and some are outlined in Sections 1 and 3 above but further explanation can be found as to whether conclusions to motivate EITM are "simply assumed." Consider the following as it pertains to formal modeling practices (Granato and Scioli, 2004: 315):

> The assumptions on which some formal modeling rests are often so at variance with empirical reality that model results are dismissed out of hand by those familiar with the facts. The problem is not just unreal assumptions, for one way to build helpful models is to begin with stylized and perhaps overly simple assumptions, test the model's predictions, and then modify the assumptions consistent with a progressively more accurate model of reality. Yet these follow-up steps are too often not taken or left unfinished, with the result being a model that does little to enhance understanding or to advance the discipline.

It is recognized also that abstract modeling is useful – indeed it is fundamental to the EITM framework. The point of departure is whether data exist to force changes in simplifying assumptions[23]:

> One justification for "theories of unreality" is that realistic models are often so complex as to be of limited value. There is merit to this defense. An important function of formal modeling is to assist in identifying crucial quantitative and qualitative effects from those that are of minimal importance. However, the drive for simplicity can be taken too far.

Moreover, it matters to respect and explain well understood empirical generalizations:

> The use of simplifying assumptions is in principle a virtue and remains so when such simplifications do no harm to overall predictive accuracy. However, this does not mean that formal modeling should proceed without regard to glaring factual contradictions in its foundational or situation-specific assumptions. Rather, formal modelers must be especially careful to make sure that they test their models in situations that go beyond the circumstances that suggested the models, for it is there that simplifying assumptions are likely to lead to difficulties.

Poor current empirical modeling practices are equally at fault and again serve as motivation. As Granato and Scioli (2004: 316–317) point out:

> … the following ratio is the subject of much attention by applied statistical analysts because it is the basis for which "theories" survive or perish:
>
> $$\frac{b}{s.e.(b)}$$
>
> This ratio is commonly referred to as a "*t*-statistic". It is the "truth" that most applied statistical analysts are concerned with, and it can be confounded by influences that shift the numerator **(*b*)** in unforeseen ways. The denominator, the standard error **[s.e.(*b*)],** also is susceptible to numerous forces that can make it artificially large or small. In either case, avoiding false rejection of the null hypothesis (Type I error) or false acceptance of the null hypothesis (Type II error) is imperative. While the concern with Type I and Type II errors should be of prime importance, that unfortunately is not usually the case. Instead, the focus is on the size of the *t*-statistic and whether one can get "significant" results.

[22] EITM Framework examples are found in Granato (2005), Granato, Lo, and Wong (2010a, 2010b, 2011).

[23] On this matter, Robert Solow (1956: 65) maintains:

> [a]ll theory depends on assumptions which are not quite true. That is what makes it theory. The art of successful theorizing is to make the inevitable simplifying assumptions in such a way that the final results are not very sensitive. A "crucial" assumption is one on which the conclusions do depend sensitively, and it is important that crucial assumptions be reasonably realistic. When the results of a theory seem to flow specifically from a special crucial assumption, then if the assumption is dubious, the results are suspect.

In a related point, Pfleiderer (2014) argues that since theoretical modeling is often undertaken to understand the implications of a given set of assumptions it is often possible for researchers to "cherry pick" assumptions to produce a desired result. It is reasonable to ask whether a theoretical model is based on assumptions that are generally consistent with what we know about the world and are capturing the most important factors.

One practice is data mining:

> The first tendency in trying to achieve "significant" results is the practice of data mining. Some political scientists put data into a statistical program with minimal theory and run regression after regression until they get either statistically significant coefficients or coefficients that they like. This search is not random and can wither away the strength of causal claims.

A related practice is creating overparameterization — by design:

> A second practice is that many studies degenerate into garbage-can regression or garbage-can likelihood renditions. By a garbage-can regression or likelihood we mean a practice whereby a researcher includes, in a haphazard fashion, a plethora of independent variables into a statistical package and gets significant results somewhere. But a link with a formal model could help in distinguishing the variables and relations that matter most from those that are ancillary and, probably, statistical artifacts. More often than not there is little or no attention paid to the numerous potential confounding factors that could corrupt statistical inferences.

A third empirical modeling practice can be characterized as having the mindset that "we are never wrong, but sometimes we are a little weak on being right:"

> The first and second practices lead to the third — statistical patching (i.e., the use of weighting procedures to adjust the standard errors [**s.e.(*b*)**] in the *t*-statistic ratio above).
>
> Statistical patches are seductive because they:
>
> have the potential to *deflate* the standard error and *inflate* the *t*-statistic, which, of course, increases the chance for statistical significance … There are elaborate ways of using error-weighting techniques to "correct" model misspecifications or to use other statistical patches that substitute for a new specification. For example, in almost any intermediate econometrics textbook one finds a section that has the Greek symbol Omega **(Ω).** This symbol is representative of the procedure whereby a researcher weights the data that are arrayed (in matrix form) so that the statistical errors, and ultimately the standard error noted above, are sometimes reduced in size and the *t*-statistic then may become significant.[24]

Clarke and Primo further assert:

> [In the 2002 EITM Report] … essays written by the [EITM Workshop] participants are hugely ironic as they make clear that the split between theory and empirical analysis is far from problematic … In the executive summary, however, this ambivalence gives way to statements such as "Significant scientific progress can be made by a synthesis of formal and empirical modeling" (page 49).

Clarke and Primo's statement misses various workshop concerns on this matter. One issue was how widespread the divide was. The 2002 EITM Report states that workshop participants, based on their professional experiences and background, believed the divide was broad:

> In their deliberations, EITM Workshop participants were in general agreement that the separation was somewhat natural and is not confined to political science. The divide exists in other social sciences, including economics, where individuals specialize in either formal or empirical analysis due to their level of mathematical background and the type and years of training the substantive area or field requires. The divide also exists in the other sciences. It was noted, for example, that epidemiology is much more comfortable with empirical modeling. The primary epidemiology journal, The American Journal of Public Health (AJPH), does not usually publish articles that have substantial formal modeling. The major funding organization for epidemiological research, NIH, tends to support very few formal modeling projects (page 5).

Adding to what was stated in Section 1 and Section 3, the 2002 EITM Report also states this natural occurrence had a negative scientific effect:

> In sum, EITM Workshop participants were in agreement that compartmentalization was not neutral in its effect. The effect is negative. It was proposed that one way to reduce the effects of compartmentalization was to separate political science into the study of domestic and international politics. Theory, data, and method would cover more general circumstances and lead to deeper understanding.[25] For the purposes of reducing the formal and empirical modeling divide, the effect of reduced compartmentalization by substantive field would encourage integration between formal and empirical analysis (EITM Report 2002: 7).

Clark and Primo also express skepticism about the need for EITM in the following passage:

> Putting aside the issue of whether empirical models are good at testing theoretical models, these justifications

[24] Note empirical modeling practices that treat patching as a robustness check can be useful:

> In principle, there is nothing wrong with knowing the Omega matrix for a particular statistical model. The trouble comes in how one uses it. Consider that Omega matrices remove or filter residual behavior. The standard error(s) produced by an Omega matrix should only serve as a check on whether inferences have been confounded to such an extent that a Type I or Type II error has been committed.

Various robustness checks on model specification and results exist (e.g., Leamer, 1983, 2010). Zellner (1984: 9–10), for example, provides other robustness tests, some of which involve the linkage of formal and empirical analysis. These tests include: 1) Studying incorrect predictions; 2) Studying implications of various equations (alter them); 3) Simulating a model's properties; 4) Pushing theories to their extreme; 5) Observing unusual historical periods; 6) Cross level inference; and 7) Experiments.

[25] Workshop participants noted: An abbreviated list of research questions that are not studied adequately because of compartmentalization are: political corruption, size of government, levels and types of taxation, economic growth and development, public debt, inflation, failed democracy, democratic stability, regime transitions, the rule of law, property and political rights, ethnic conflict, coups and revolutions, and terrorism (EITM Report: 7).

are vague regarding their very premises: how previous practices harm the discipline, and how better theory testing improves the discipline. There is no evidence that continuing our current practices might “delay, or worse, derail the momentum generated over the past 40 years” (Granato and Scioli, 2004: 313) (Clarke and Primo, 2012: 139).

Again, and contrary to Clarke and Primo's assertions, the 2002 EITM Report provides discussion of current methodological practices. It is a key discussion point in the 2002 NSF EITM Workshop. When it came to sources of the problem (and recall from Section 1), workshop participants focused on:

> Differences between formal and empirical approaches occur in intellectual outlook, skills, training, and research focus. In terms of outlook, formal modelers typically emphasize, in minute detail, linkages between concepts, whereas empirical modelers do not want to spend their research time parsing through minute details that may not add to their understanding. Formal modeling also requires analytical, logical, and mathematical modeling skills, while empirical modeling is inductive and, therefore, places emphasis on descriptive and statistical skills. Workshop participants noted that the intellectual investment needed for formal modeling is greater; it requires more mathematical knowledge than does empirical modeling to analyze a problem of interest. Training priorities differ as well. Empirical modelers devote their energies to data collection, measurement, and statistical matters, while formal modelers center on mathematical rigor (EITM Report 2002: 5).

As stated previously in Sections 1 and 3 above, the effects were not considered neutral. Creating better ways for testing was not the issue. Instead, current practices were due to self-reinforcing resistance to improvement (EITM Report 2002: 5):

> These differences in outlook, skills, and training are reflected in distinct research practices and outcomes. For empirical modelers, model failures lead to emphasis on more statistical training or more sophisticated uses of statistics – usually to “patch over” – a model failure (See Appendix A). Formal modelers, on the other hand, deal with model controversies by considering alternative mathematical formulations but this is usually done piecemeal. The basic framework, such as expected utility, usually remains in place. The one similarity, however, between these two approaches is that both formal and empirical modelers tend to remain tied to their particular technique despite the warning signals evidenced in model breakdown.

As to evidence of the problems with current practice and their consequences, we can either select past studies or – as an alternative – reference scholars who specialize in particular areas and in different disciplines. We choose the latter option and use Granato and Scioli's (2004) reference to Akerlof (2002) and Achen (2002) in their criticisms of formal modeling and empirical modeling respectively:

> This conflict between realism and analytical tractability is not new or only a problem in the discipline of political science. Economics is instructive in this regard. In the late 1960s and early 1970s there was a revolution in macroeconomic research, which put great emphasis on the microfoundations of macroeconomic outcomes. Yet, as George Akerlof recently noted:
>
> [T]he behavioral assumptions were so primitive that the model faced extreme difficulty in accounting for at least six macroeconomic phenomena. In some cases logical consistency with key assumptions of the new classical model led to outright denials of the phenomena in question; in other cases, the explanations offered were merely tortuous (Granato and Scioli, 2004: 315).[26]

As to the scientific problems with specific empirical practices Granato and Scioli (2004) state:

> If one were to summarize the problem here, one would conclude that the intellectual drift from the virtues of empirical practices means that statistical technique has come to dominate the practices used to help identify causal linkages. But statistical technique alone cannot test generalizations of observed political behavior. Once again, the solution is to find ways to link statistical techniques with formal theory:
>
> Traditionally we have tried to do both with informal assumptions about the right list of control variables, linearity assumptions, distributional assumptions, and a host of other assumptions, followed by a significance test on a coefficient. But since all the assumptions are somewhat doubtful and largely untested, so are the estimators and the conclusions. The depressing consequence is that at present we have very little useful empirical work with which to guide formal theory. The behavioral work too often ignores formal theory. That might not be so bad if it did its job well. But it produces few reliable empirical generalizations because its tests are rarely sharp or persuasive. Thus, empirical findings accumulate but do not cumulate (Granato and Scioli, 2004: 317).[27]

4.2. EITM and the Hypothetico-Deductive (H-D) method

Clarke and Primo attempt to establish a link between EITM and the H-D method. They describe H-D as follows:

> The H-D approach comprises the following:
>
> - a hypothesis H set up for testing or examination;
> - an observation sentence O implied by H along with theoretical background statements, mixed statements, boundary conditions, etc.; and,
> - an experiment or examination of the world where we observe either O or ~O.[28]

26 Akerlof (2002: 412).

27 This latter quote is from Achen (2002: 445).

28 Clarke and Primo reference Kyburg (1988: 65) for this definition.

> If we observe ~O, then we have refuted H. If we observe O, then we have confirmed H or, at the very least, failed to refute H. Less formally, "Theory implies prediction (basic sentence, or observation sentence); if prediction is false, theory is falsified; if sufficiently many predictions are true, theory is confirmed." (Clarke and Primo, 2007: 744).

Clarke and Primo "trace the evolution of the initiative" and argue they can "demonstrate" EITM bridges the gap between formal and empirical analysis "using H-D" (page 48). They argue a connection for this assertion can be found in Granato and Scioli (2004: 315):

> Granato and Scioli (2004) are quite specific on the role that H-D should play in political science and they elaborate their ideal world:
>
> In an ideal world, where there is unification in approach, political science research should have the following components: 1) theory (informed by field work, or a "puzzle"); 2) a model identifying causal linkages; 3) deductions and hypotheses; 4) measurement and research design; and 5) data collection and analysis.

What Clarke and Primo fail to mention is this quote builds on a more general point about research design competence and overall proposal competitiveness for NSF proposals (i.e., Motivations 1 and 3 in Section 1). Specifically, these same points were discussed in the 2001 EITM Workshop. The 2002 EITM Report summarizes the issues of basic research design construction:

> In an ideal world, political scientists should be educated to do research that incorporates five major components: 1) theory (informed by field work or some "puzzle"); 2) a mathematical model identifying causal linkages; 3) deductions and hypotheses; 4) measurement and research design; and 5) data collection and statistics. However, one or more of these components often is absent in political science research and as argued by the EITM Workshop participants, the quality of formal and empirical modeling in political science is substandard (page 7).

The question we have is this. Are students not to be exposed to these elements in a research design (e.g., scope and methods) course? The idea that students should not be trained to know the basics of deductive reasoning, hypothesis formation, data collection and statistics (analysis) strikes us as impairing student development with harmful future consequences for any scientific discipline. We do not think Clarke and Primo believe this either. But, because they failed to consider the relevant documents – and also never discuss the EITM framework – they have no basis for evaluation of how EITM is implemented and defined at NSF. The appropriate "test" for their assertions is to evaluate the EITM framework.

Clarke and Primo's (2007) solution to their description of the H-D matter is to "abandon the practices of the hypothetico-deductivism" (Clarke and Primo, 2007: 748). They suggest the following rules to integrate models and data. We state Clarke and Primo's rules below and relate them to the EITM framework.

Clarke and Primo's rules are as follows (Clarke and Primo, 2007: 748–749):

> 1. Be clear about the purpose(s) your model is intended to serve.
> 2. Abandon the goal of "model testing" as currently practiced. "Model testing" implies using statistical analysis to determine the truth or falsity (or any of the synonyms that political scientists use, such as "supported," "confirmed," "verified," or "validated") of a model, but as discussed earlier, the truth or falsity of a model is not the question. Rather, the point is demonstrating that the model is useful in a particular way … [a reason one can forego] data analysis in … [a] structural model is that the field possesses … a number of strong empirical generalizations. In a research area with fewer such generalizations, data analysis would be required to make a compelling case.
> 3. Include a data analysis only when the purpose(s) of your model is served by it. Not all models require an accompanying data analysis … A researcher should be clear about how the data analysis supports the purpose of the model, and if it does not support the purpose of the model, leave it out.
> 4. Treat data analysis as more than an endpoint. On those occasions where models and data are integrated, too often the model is carefully developed over the first nine-tenths of the paper while an inconsequential data analysis is tacked on as the final one-tenth of the paper, no doubt to appease reviewers. Seeing data analysis simply as an endpoint is an unfortunate consequence of a focus on model testing … True integration of models and data is not easy. Showing that a model is similar to the world in a particular way for a particular purpose often has description, as opposed to inference, as its goal, and to some political scientists, description is a dirty word evoking a theoretical accounts devoid of conceptual bite. When guided by theory, however, description becomes a powerful tool both for assessing the usefulness of a model and for opening new avenues for theoretical exploration.

Now, recall the EITM framework is as follows:

Step 1. Relate and Unify Theoretical Concepts and Applied Statistical Concepts

Step 2. Develop Behavioral (Formal) and Applied Statistical Analogues

Step 3. Unify and Evaluate the Analogues.

Can one decipher H-D from the EITM framework? How one "tests" a model is of crucial importance and the EITM framework does not preclude any type of testing so long as there is an explicit tie between the formal and empirical analogues.

Clarke and Primo do emphasize that "theoretical models can be used to explain findings or generalizations produced by empirical models. Empirical models, on the other hand, cannot provide explanations" (Clarke and Primo, 2012:

137). But, this is consistent the EITM framework. As Wesley Salmon (1988: 6) notes, a theory is a collection of models or "a set of models." The EITM framework demonstrates the term "models" should include not only theoretical models but also empirical models since both possess concepts and analogues for the concepts.

In Clarke and Primo's view theoretical models are not tested with data; they are tested with models of data, which are far from secure. We do not view testing theoretical models with models of data as problematic because the models of data *can* represent phenomena of interest. *This is precisely the point for using analogues*. Therefore, the EITM framework – and analogue development – provides a link between theoretical and statistical concepts. Once the conceptual link is built, we have the opportunity to determine what kind of empirical and formal model analogues best represent the given theoretical and empirical concepts. An important measure of scientific progress is the improvement in analogue development for our concepts of interest.

The H-D method, by way of contrast, begins with a theory about how things work and derives testable hypotheses from it and its focus is to use empirics to test the hypotheses that then support or discredit a theory. In short, the H-D method concentrates on the relation between hypotheses and empirical tests, but is not necessarily about unification – a transparent and direct link between theory and test.

The EITM framework does allow for the logical "truth-preserving nature" of deduction used by theoretical models (e.g., Arrow, 1963), but a logically true conclusion does not mean it fits the facts or serve as an "explanation." Robert Lucas (1988) argues "the role of theory is not to catalogue the obvious, but to help us to sort out effects that are crucial quantitatively, from those that can be set aside (page 13)." But, the matter of ascertaining "crucial quantitative importance" requires the theory fits the facts *in addition* to achieving substantive significance. Anthony Downs' (1957) classic model of voting is an example of seeking both consistency between logic and empirical truths. Downs' model of voting predicts a unique low turnout; however, we find a reasonable number of people go to the polls in the real world. Therefore, the "logical truths" of theoretical models do not guarantee they are empirical truths. Empirical analysis – model testing – assist in developing and revising the explanation.

However, the use of the EITM framework does not mean the disconfirmation of a theoretical model signifies the failure of a theory. The reason is that various theoretical models can be developed from the same theory. In other words, a theoretical model simply reflects a specific dimension of a theory. Clarke and Primo (2012: 50) are correct to criticize Granato and Scioli (2004: 314) for being "ruthless" in their assertion that we can discard a theory based on a limited sample of predictive failure. The issue in model rejection is more complicated: the dialogue requires a far broader testing regimen (e.g., alternative methods, data, and the like) as well as specificity in just what is discarded.

A case in point is rational choice theory. The supposed failure of a rational choice model to account for turnout does not mean the failure of rational choice theory. Other theoretical models derived from rational choice theory might help explain and predict other aspects of human behavior.[29] Furthermore, even though scholars use the same theoretical model to explain the same behavior, they might have different definitions of – or perspectives – concerning the components of a theoretical model. Again, note the dialogue with an empirical component improves the explanation.

Under rational choice the turnout decision can be characterized by a decision calculus balancing four factors (i.e., P, B, C, and D).[30] A citizen's turnout decision can be expressed as: $R = (BP) - C + D$, where R represents the expected utility of voting. Accordingly, if $R > 0$, the citizen goes to the polls. On the other hand, if $R \leq 0$, a citizen abstains from voting.

The turnout paradox – and the failure of this particular model to fit the facts – led scholars to devise different arguments about these four factors. In terms of the cost of voting, some argue that the opportunity and transportation costs of voting are overblown (Aldrich, 1993; Palfrey and Rosenthal, 1985), whereas others contend voting costs are significant (Converse, 1964; Brians and Grofman, 1999). There is also debate on the probability that one's vote influences the outcome (P). Some assume P is a fixed quantity, whereas others assume P is a parameter arising endogenously from the strategic interaction of citizens.

The point here is the same theoretical model, with different assumptions about the components, generates different theoretical explanations and predictions even though they all satisfy the truth-preserving nature of deduction and are logically true. However, when rival explanations are in play, how can we judge which one is better? Accordingly, it is reasonable to argue the theoretical hypotheses derived from a theoretical model should be tested.[31]

Is it inappropriate then to use the words such as "test," "support," or similar phrases to describe the evaluation process of the theoretical models? We think wording here is less a concern than the actual activity. Again, the idea of EITM is to develop better "connections" between theory and empirics to improve understanding about the relation between X and Y. In the EITM framework it is accepted that theory can provide clarity to data, but we think the reverse is also possible. We argue the feedback from empirical

[29] For example, political scientists developed a number of theoretical models (based on rational choice) to explain turnout (e.g., Downs, 1957), collective action (e.g., Olson, 1965), legislative behavior (e.g., Riker, 1958), and electoral competition (e.g., Stokes, 1963).

[30] P refers to the probability that one's vote influences the outcome; B refers to the benefits a voter receives from seeing her preferred candidate win; C refers to the cost of voting; finally, D refers to the utility one receives as a direct consequence of casting a ballot.

[31] A case in point is when there exists conflicting results in empirical studies, theoretical models provide guidance explaining how one result differs from another. Would the data collected for one study satisfy one set of assumptions while the data for another study satisfies another set of assumptions? For example, in the literature of foreign direct investment (FDI), the empirical results estimated based on the data from developing countries can be very different from the results based on the data in developed countries (Blonigen and Wang, 2005). This circumstance suggests that researchers should carefully impose the assumptions or characteristics of a theoretical model when they are studying different groups of countries.

testing – pursuant to a dialogue – can help revise and further develop theoretical models – even establish new connections between theory and empirical findings. Put differently, better theory can be motivated by previous theoretical assumptions with the assistance of empirical results. For social science questions this dialogue between theory and empirics (tests) (deductive and inductive reasoning) sustains a deeper or a broader exploration.

From Clarke and Primo's rules above it is also unclear how they would sort out *the usefulness* of what they see *as useful models*. Would not data and testing enter into this process? Even undertaking logical exercises modelers at some point would need to know how much their argued for factor or factors matter. Moreover, a problem with Clarke and Primo's rules is the failure to consider how *empirical models cannot be divorced from empirical practice*. Destructive empirical practices have been outlined above and this is something Clarke and Primo are silent about.

As a final point we are struck by Clarke and Primo's focus is on the past and how they try to fit the EITM initiative into a box. Their criticisms about "testing" conjure up long ago debates including John Maynard Keynes (1939) critique of econometric methods and their usefulness. Then as now formal and empirical tools continue their forward progress but it is a mistake to think this progress in tool development will not foster tighter linkages between theories and tests. Unlike Clarke and Primo's rules, the EITM framework is explicit about creating a dialogue between theory and test. This enhanced dialogue allows us to improve upon our current assumptions that often are short-cuts for the current state of data and formal and empirical tools.

5. Ensuring a dialogue between theories and tests

While it is important to do pure theory and pure empirical work, we have argued a motivation for EITM is that current practices (in some cases) have now reached a point of diminishing returns if not outright harm to the scientific process. Moreover, we think rules proffered by Clarke and Primo, because they fail to engage harmful empirical practices discussed earlier, perpetuate harmful aspects of the methodological status quo. This includes detrimental testing practices, but we think they are more likely to occur when researchers engage in data mining, overparameterization, and statistical patching.

Clarke and Primo's arguments diminishing the use of empirics and testing forgoes useful and feedback and dialogue from empirical observations in refining theory which can foster cumulation of knowledge. The dilemmas where theory is ahead of data or data are ahead of theory can be dealt with more effectively by employing a form of methodological unification including the EITM framework.

An important issue in the dialogue is the whether the tests involved predictions or fitting facts. The distinction between the two is difficult and often blurred. Often they work in concert and their difference distracts from creating a linkage with feedback to create a dialogue between model and test:

> Moreover, this linkage and evaluation provides a clearer (i.e., falsifiable) interpretation of the results because the model's mechanisms are explicit. The specified linkage shows potential sources of inaccuracies and model failure. Further, an inductive and deductive dialogue is created between the data and the technique(s) where new theoretical mechanisms and/or new analogues can be used (Granato, 2005: 12).

5.1. *Elinor Ostrom's research on social dilemmas*

An example of the dialogue we are describing and what is consistent with the EITM initiative is Elinor Ostrom's (2010) work, *Revising Theory in Light of Experimental Findings*. In her article Ostrom explains how the empirical findings from experiments move the theory forward with regard to the study of social dilemmas. Game theory provides a clear theoretical assumption that can be tested not only by secondary data, but carefully designed experiments. Furthermore, making changes in the experimental design allows examination of game theoretic predictions and further develops the game theoretic model. As Ostrom (2010: 69) states: "behavior in social dilemma experiments varies dramatically from being consistent with game theory predictions to being inconsistent, depending on the design of the experiment." Ostrom (2010) shows more specific examples of how feedback from experimental tests supports revision of the conventional self-regarding model in social dilemmas.[32]

Scholars find diverse experimental environments bring an immense variety of outcomes related to social dilemmas rather than conforming to theoretical predictions simply from a model of the individuals who maximize their own short-term payoffs (Ostrom, 2010: 70). For instance, Ostrom and Walker (1991) allow face-to-face communication in the experiment, which leads to a substantial reduction in subjects' overharvesting. David Sally (1995), in his meta-analysis of 35 years of published experiments related to prisoner's dilemma games, finds communication among the subjects significantly changes their degree of cooperation in repeated experiments. Schmitt et al. (2000) find the influence of communication depends on decision maker involvement in the discussion *after* a common-pool resources (CPRs) experiment *and* also letting the subjects in the subgroup be party to the discussion.[33]

The experimental results – *the empirical tests* – also challenge the presumption that only externally imposed regulations can make people overcome social dilemmas.

[32] In the language of the EITM framework, Ostrom is working with theoretical concepts related to decision making, learning, and social interaction. The empirical concept centers on prediction. The formal and empirical analogues include game theory and simple numerical and qualitative outcomes (See Section 2).

[33] Participants in the CPR experiment are asked to make decisions about the investment of an endowment of tokens between two markets. Market 1 yields a certain, private return while the return from tokens invested in Market 2 is dependent on both individual and aggregate token investments in that market. For more detailed information about the CPR experiment, please refer to Ostrom et al. (1994).

Lopez et al. (2009) find informal sanctions and subject knowledge of the group in a framed field experiment is more effective than external rules. Similarly, Ostrom et al. (1992) find subjects punish those who keep over-harvesting if opportunities for engaging in costly punishment were given in CPR lab experiments.[34]

In sum, these empirical findings and tests from diverse experiments indicate the experimental environment and context can change behavior related to social dilemmas, and this is inconsistent with the predictions based on the self-regarding model. It is inadvisable to use the conventional model for predicting outcomes in experiments. Moving forward we need to consider changing the behavioral theory.

What do various empirical tests offer as new information? Numerous studies of social dilemmas have proposed a variety of alternative models of individual behavior in which expanding the potential factors individuals consider during the decision process (Camerer and Fehr, 2006; Cox et al., 2007). Individuals in social dilemmas tend to value returns to others rather than solely seeking their own immediate benefits. And how much individuals value returns to others depends on who the others are, their joint history, and information available about past behavior. Individuals apply norms of behavior in a variety of settings besides material interests (Crawford and Ostrom, 1995).

It is unsurprising when different individuals behave differently in distinct experiments, but what is also clear is the self-regarding assumption is inconsistent across experiments and field settings. In addition, individuals do learn norms of behavior and are affected by these norms. A different theoretical model would be based on a behavioral theory of the self-regarding individual who possesses "other-regarding" preferences and internal norms. This results in a higher level of cooperation than predicted by the conventional theory (Ostrom, 2010: 70).

An additional point to note is the role of "context" and the importance qualitative attributes such as this play in model and test revision and improvement. First, context refers to the experimental environment designed by the researcher to test theory. Second, it also represents the broader context outside affecting preferences and actions. However, this is not the end of the story. As Ostrom (2010: 70) asserts:

> when a careful design repeatedly does not produce the predicted outcomes, we have to ask, is it the design? Or, is the theory of the individual being used? I am arguing that we need to assume a more complex theory of the individual when we study social dilemmas.

Indeed, thinking further, we find individuals are likely to revert to their own immediate benefits (self-regarding) when there is no chance to know those with whom they interact, no information available about their past behavior, or no communication in the experiments. Alternatively, individuals tend to cooperate more if trust and reciprocity are well developed because of the two context aspects, and the role repetition plays in reputation acquisition.[35]

Theory, then, can be modified to take into account the role of reputation, trust, and reciprocity affecting outcomes in repeated social dilemmas (Ostrom, 2010: 71). It is possible that when facing social dilemmas, a cooperator will not be a sucker who contributes while others continue to free ride because individuals can be trustworthy reciprocators. This can help us understand and extend experimental findings related to the governance of natural resources.

Ostrom (2010) provides an important example of testing and consistency evaluation. She explains how empirical findings from experiments assist in testing the game theoretic model of the self-regarding individual with no cooperation and how the testing feedback works to revise the theory for social dilemmas: from self-regarding to other-regarding, adding context, reputation, trust, and reciprocity.

6. Limitations of EITM

The EITM framework possesses limitations. These shortcomings center on *observational equivalence* and *analogue development*. To begin, observational equivalence is related to identification.[36] Also, recall reduced form estimates fail to provide structural parameters and this requires use of model or parameter restrictions so identification is achieved. Observational equivalence occurs when two or more rival models provide statistically indistinguishable reduced form results. Moreover, observational equivalence can occur even if the respective models are identified. An important paper on this issue is by Thomas Sargent (1976). In his review of this issue Patrick Minford (1992) summarizes Sargent's (1976) results as follows:

> … models may be fully identified; that is, the parameters of each may be individually retrieved by estimation of the full model (i.e. subject to all its restrictions). However, there is a useful potential connection with the concept of identification. If two models can be 'nested' in a more general model (usually a linear combination of the two), then, provided the coefficients of each model

[34] Challenges can be found in other motivation sources. Findings from 40 laboratory and field experiments conducted around the world challenge the theoretical assumption that people only focus on material interests for self (Bowles, 2008).

[35] Increasing evidence from neuroeconomics is that some individuals gain real pleasure from norms of behavior such as trustworthy action (McCabe and Smith, 2001; Fehr et al., 2005), which is consistent with Crawford and Ostrom's (1995) inclusion of the concept of norms in the preference functions of individuals making a cooperative move. As Ostrom (2010) discusses, norms affect individual behavior because people appear to treat positive actions with positive responses and vice versa. Gaining a reputation for being trustworthy and reciprocating cooperation help enhance cooperation and to achieve substantial long-run interests and increase individual net benefits substantially.

[36] Recall the EITM framework builds on the Cowles Commission's contributions. However, methodological unification is broader than standard econometric tools. We fully expect innovations beyond what we know today to be consistent with the EITM framework and its aim to create a dialogue between theory and test.

> can be identified in this general model, it is possible to test for their significance and accordingly that of each model. In this situation, if (and only if) the coefficients cannot be identified, the models will be 'observationally equivalent' (page 425)

The good news is this challenge in distinguishing between rival models can be narrow – occurring in one dependent variable, but not in other dependent variables – but the bad news is its existence is still a problem. Potential solutions do exist but none are generalizable. They are for a specific case. These solutions include either imposing theoretically justified exclusion restrictions or identifying regime shifts that can yield theoretically distinct predictions. A combination of both is also possible, but this would depend on the specific set of models and data.

With regard to analogue development, two technical challenges emerge. One technical challenge is in developing analogues. Unlike the natural sciences, social sciences study human subjects possessing expectations affecting their current behavior. This "dynamic" creates moving targets for many social science questions. How to improve upon current analogues for distinctly human behavioral traits (e.g., norms, reputation, expectations, learning) is a key future hurdle to achieving scientific cumulation.

A second technical challenge relates to the framework's emphasis on *parameters* as a building block for *ex-post* and *ex-ante* prediction. It is almost impossible to capture all parameters in complex political, social, and economic systems. However, the EITM framework is useful since it helps researchers open the "black box" relating different theoretical parameters to the estimated coefficients in an empirical model. *A more general point concerns the EITM framework's focus on parameters separates variables that aid in fundamental prediction from other variables considered "causal" but are of minor predictive importance* (See Lucas 1988).

7. An application to economic voting

Past studies of economic voting normally assume voters hold government responsible for changes in their personal financial situations. Under this assumption, theoretical specifications for the cross-sectional estimates of these studies are similar to the time-series estimates in macro-level studies. This similarity in specifications has led researchers to expect that the individual-level studies produce findings consistent with the aggregate-level ones. At the aggregate level, studies find significant effects of economic performance on election outcomes. It turns out many of the micro-level research findings reveal parameter instability as well as variation in economic vote magnitude. These cross-sectional studies fail to uncover an individual-level basis for the macro-level relation between economic circumstances and vote choice.

Some studies resort to purely statistical fixes in the hope these empirical "tools" solve the conundrum. Among them are studies using instrumental variables to estimate the (reciprocal) links between presidential approval and party competence (Kinder and Kiewiet, 1979), sophisticated maximum likelihood logit model (Fiorina, 1978), and a combination of OLS, probit analysis, and two-stage least squares (Kinder and Kiewiet, 1981).[37]

7.1. A theoretical solution: opening the way to EITM

A theoretical explanation for the conflicting empirical results has been offered by Gerald Kramer (1983). He argues economic voting has both a governmental-induced component and an exogenous component – the latter determined by the life-cycle and other factors that are beyond governmental control. Voters respond not to changes in their real income as a whole, but instead only to the portion of the change that is attributable to government policy. This explains, to a large extent, why pure empirical approaches to estimating the effect of personal financial conditions on vote choice fail to produce satisfactory findings.

Kramer's assumption contains important formal and empirical implications. The behavioral relation to be estimated involves only the government-induced component, but since both the government and non-government components are not observable to the voter, we have to deal with the "noisy" version of the variable which poses complicated estimation problems (i.e., a signal extraction problem). More significantly, Kramer's example illustrates the importance of the linkage between theory and empirical analysis in social science research.

In the case of economic voting research, this linkage means a careful treatment of conceptual issues and their implications for empirical analysis. Among the issues to consider are the dynamic formulation of expectations, the separation of policy and non-policy effects, and ideas about public sophistication and political accountability (See Kiewiet and Rivers, 1984). *The connection between these issues and the EITM framework is revealed in the use of concepts and analogues pertaining to expectations, uncertainty, and measurement error.*

7.2. EITM and economic voting

A number of studies extend Kramer's theoretical innovation to improve understanding of economic voting. Important and noted work can be found in Alberto Alesina and Howard Rosenthal (1995). Their EITM approach has a strong interdisciplinary flavor, drawing on the work of Milton Friedman (1957).[38]

[37] Applied statistical tools can lack power in disentangling conceptually distinct effects on a dependent variable. This is noteworthy since the traditional applied statistical view of measurement error is that it creates parameter bias, with the typical remedy requiring the use of various estimation techniques.

[38] Friedman (1957) contributes to the study of the "signal extraction" problems by linking specific empirical coefficients to his behavioral model instead of treating his research question as a pure measurement error problem requiring only applied statistical solutions. Specifically, he merges "error in variables" regression with formal models of expectations and uncertainty concerning permanent and temporary changes in income.

7.2.1. The relation between expectations, uncertainty, and measurement error

Alesina and Rosenthal relate the behavioral concepts of expectations and uncertainty to a measurement error problem. They develop a formal model based on the – expectations-augmented – Lucas aggregate supply curve (Lucas, 1973).

$$y_t = \bar{y} + \gamma\left(\pi_t - \pi_t^e\right) + \varepsilon_t. \tag{1}$$

In Equation (1), economic growth is a function of inflation expectations (a government policy component) and uncertainty about incumbent competence (a non-policy component). The variable, y_t, is the rate of economic growth at time t, and $\bar{y}$ is the natural (average) rate of growth. The component in the parenthesis denotes government policy: π_t is the actual inflation rate at time t, and π_t^e is the expected inflation rate at time t formed at time $t-1$. The model implies growth is positive (negative) when the actual inflation rate is higher (lower) than the public's expected inflation rate.

Voters' uncertainty about competence of the incumbent is captured by the error term, ε_t, which represents the unobserved shock to economic growth. The analogue in Equation (2) formalizes the concept of uncertainty as a "signal extraction" or measurement error problem. Specifically, ε_t is broken down into two component parts: η_t is competence attributed to the incumbent administration relative to the other party, and ξ_t is the shock to growth beyond administration control – good or bad "luck." Both η_t and ξ_t are mean zero with variances σ_η^2 and σ_ε^2 respectively.

$$\varepsilon_t = \eta_t + \xi_t. \tag{2}$$

The model depicts myopic voters regarding the incumbent component. Specifically, uncertainty about incumbent competence follows a first-order moving average process (MA(1)). In Equation (3), μ_t is incumbent competence for the current period (μ_t is iid, $(0, \sigma_\mu^2)$, μ_{t-1} is competence for the last period, and ρ denotes strength of the prior period's effect. Evaluating competence is based on performance in the prior period: it is fully realized at time t, and either fully or partially realized at lag time $t-1$.

$$\eta_t = \mu_t + \rho\mu_{t-1}, 0 < \rho \leq 1. \tag{3}$$

The model determines competence in the following ways. If the voter predicts inflation with no systematic error (that is, $\pi_t = \pi_t^e$, hence, $\gamma(\pi_t - \pi_t^e) = 0$), then the economic growth rate deviation from the natural growth rate is attributed solely to the incumbent's competence (η_t) and the stochastic economic shock (ξ_t):

$$\varepsilon_t = \eta_t + \xi_t = y_t - \bar{y}. \tag{4}$$

But, voters are assumed to have some uncertainty about competence. In equation (4), when $\varepsilon_t = \eta_t + \xi_t > 0$, the actual economic growth rate is greater than the natural growth rate. This implies that the voter will determine whether this above-average economic growth is due to competence (η_t), economic shocks (ξ_t) or both.

Now, suppose voters make forecasts based on the belief that competence can persist and so they give greater or lesser weight to competence over time. This behavioral concept can be formalized using conditional expectations as an analogue for the optimal forecast of competence (η_{t+1}):

$$\begin{aligned} E_t(\eta_{t+1}) &= E_t(\mu_{t+1}) + \rho E(\mu_t|\mu_t + \xi_t) \\ &= \rho E(\mu_t|\mu_t + \xi_t) \\ &= \rho E(\mu_t|y_t - \bar{y} - \rho\mu_{t-1}), \end{aligned} \tag{5}$$

where $\mu_t + \xi_t = y_t - \bar{y} - \rho\mu_{t-1}$, and $E_t(\mu_{t+1}) = E(\mu_{t+1}|\mu_t + \xi_t) = 0$.

In Equation (5), the rational voter can observe the composite component of competence and economic shock ($\mu_t+\xi_t$) based on the observable variables, y_t, $\bar{y}$, and μ_{t-1} which are available at time t. Competence, η_{t+1}, can therefore be forecasted by predicting μ_{t+1} and μ_t.

7.2.2. Unifying and evaluating the analogues[39]

From (5) it can be shown that the behavioral analogues for expectations and uncertainty create a link to the applied statistical analogue – measurement error:

$$\begin{aligned} E(\eta_{t+1}) &= \rho E(\mu_t|\mu_t + \xi_t) = \rho E(\mu_t|y_t - \bar{y} - \rho\mu_{t-1}) \\ &= \rho \frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma_\varepsilon^2} \cdot (y_t - \bar{y} - \rho\mu_{t-1}), \end{aligned} \tag{6}$$

where: $0 < \frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma_\varepsilon^2} \leq 1$.

Equation (6) is the competence model expressed as an error-in-variables regression. It illustrates the linkage of the behavioral analogue for expectation and uncertainty, and the applied statistical analogue of measurement error. The expected value of competence at time $t+1$, $E(\eta_{t+1})$, is positively correlated with economic growth net of natural growth, $y_t - \bar{y}$, and the proportion of competence that does not carry over to the next term, $\rho\mu_{t-1}$. The signal-noise ratio, $\frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma_\varepsilon^2}$, is the proportion of competence the voter observes and interprets and thus can be called the "competency signal" (Duch and Stevenson, 2010). Its interpretation is two-fold:

1. When $\sigma_\varepsilon^2 \to 0$, $\frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma_\varepsilon^2} \to 1$ (The voter interprets the economic shock variability as primarily a function of incumbent competence.).
2. When σ_ε^2 increases, $\frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma_\varepsilon^2}$ decreases (The voter recognizes the economic shock variable is due to random chance and assigns less weight to competence.).

7.2.3. Leveraging EITM: cumulation

The application here and the EITM framework does not mean the theory is correct. Cumulation over time can ultimately be upended by alternative theories that lay far outside a current literature. A new cumulative process would then be built around the new and novel theory. In acknowledging this reality, we make the weaker claim that

[39] See Granato et al. (2010b: 23–26) for a description of the tools used to construct these analogues.

methodological unification fosters a coherent evolving process where successive research builds on prior research and where the linkage between theory and test directs successive research to alternative measures, theoretical assumptions, and new formal and empirical analogues (Granato et al., 2010a: 784).[40] More to the point, using the EITM framework supports cumulation because the formal model, informed by various empirical tests (i.e., fitting facts, predictive accuracy), can reveal how the covariates relate to each other (i.e., understanding the inner-workings of the system). The other strength is these covariates can be investigated ex-ante and prior to doing any testing.[41]

Fig. 1 summarizes an example of selected articles originating with Kramer's (1983) effort linking formal and empirical analysis. The examples highlight the transparency between theory and test not only deepens the level of understanding through behavioral representations that improve upon simple socio-economic categorization, but also broadens researchers' ability in modeling voter behavior in different environmental and information contexts. Corresponding with these changes is a focus – as we mention above – on alternative measures, theoretical assumptions, and new formal and empirical analogues.

Since not all macroeconomic policies induce politically relevant permanent changes in output, it is reasonable to assume that voters may reward incumbents for permanent growth, but punish them for less desirable cyclical growth. Uncertainty can be tested differently by examining the competence analogue of (η_t) with alternative measures.

Because the analogue is part of the aggregate supply (AS) shock $(\varepsilon_t = \eta_t + \xi_t)$, competence (η_t) can be defined as the incumbent's ability to promote economic growth via the AS function. We can further assume voters reward capacity building AS policy for generating long-term economic growth, but otherwise punish policy that causes a short-term shift in aggregate demand (AD) with undesirable inflationary effects.

Motoshi Suzuki and Henry Chappell (1996) make use of similar arguments on permanent and temporary changes in economic growth to evaluate the competence Equation (6) by *using different measures for uncertainty*. Specifically, the authors replace the restrictive MA(1) process in the original model with flexible empirical specifications, and apply advanced time series techniques to decompose real GNP into permanent and cyclical income components. The authors then jointly estimate a three-equation system for shares of the two-party vote for presidential, Senate and House elections. Their results shed new light on voting behavior by showing that voters are more sensitive to permanent than cyclical economic growth. The policy implication is that voters' preference for permanent growth would encourage politicians to adopt policies that generate long-term rather than short-term growth.

Raymond Duch and Randy Stevenson's (2010) study applies the competence model of Alesina and Rosenthal but now have *voter decisions based on both global and domestic economic outcomes*. Using the signal-noise ratio $(\frac{\sigma_\mu^2}{\sigma_\mu^2+\sigma_\varepsilon^2})$, the expectations-uncertainty-measurement error analogue in (6), the authors provide a “competence signal” explanation for the cross-national and dynamic variations in the magnitude of economic vote.

The authors test their proposition with micro-level data from a six-nation survey. Their findings show voters not only have a sense of the total variation in economic shocks (i.e., σ_μ^2 and σ_ε^2) but they also distinguish between the relative contributions of the different components – domestic and global – of the total variation. When $(\sigma_\mu^2$ is high (or low) relative to (σ_ε^2) it suggests a high (or low) competency signal. The micro-level results are further confirmed by aggregate-level data of macro-economic time-series of 19 countries. The study further demonstrates open economies, which experience large exogenous economic shocks relative to government-induced ones (i.e., $\sigma_\mu^2 < \sigma_\varepsilon^2$), exhibit a smaller economic vote than countries that are less dependent on global trade.

Duch and Stevenson (2008) also apply their model to the attribution of political responsibility. The authors assume voters distinguish between “electorally dependent decision makers” (mainly political office holders), and those who are “non-electorally dependent” ones such as bureaucrats, business firms, interest groups, foreign leaders, the WTO, and the like. In the same way as Alesina and Rosenthal, their model uses “signal extraction” which is based on government competence and an exogenous shock to economic growth. But, their model differs in that they *assign competency shocks to the governmental decision makers and exogenous shocks to the non-governmental actors*. The specification is useful in examining how electoral outcomes are influenced by the arrangement of domestic political institutions, the ambiguity of policy responsibility, and the influence of the global economy.

In their economic voting study on Latin America, Isabella Alcañiz and Timothy Hellwig (2011) build on both Alesina and Rosenthal and Duch and Stevenson. Following Alesina and Rosenthal, their analysis treats government competency as the basis for developing expectations about political responsibility. Alcañiz and Hellwig *assign responsibility to economic policy decision makers who are electorally dependent from those who are not*. The findings from 17 Latin American countries indicate that, instead of punishing political office holders for poor economic outcomes, voters assign blame to international and private-sector actors.

[40] Is it possible to have a coherent evolving cumulative process using pure formal or applied statistical approaches? The answer is yes, but as we have outlined in Sections 1, 3, and 4 there are specific practices that are noncumulative.

[41] On this latter issue, Granato et al. (2010a) point out that:

> Even scholars who are sensitive to establishing robustness in their applied statistical results find the available tools inadequate when used in isolation. For example, augmenting applied statistical tests with Extreme Bounds Analysis (EBA; Leamer, 1983) provides a check on parameter stability, but the test is performed ex-post and therefore does not allow for ex-ante prediction … This should not be surprising when one considers the effects of previously unspecified covariates in this procedure. Each time an applied statistical model is respecified, the entire model is subject to change. But without a priori use of equilibrium conditions (e.g., stability conditions) in a formal model, the parameter “changes” in a procedure such as EBA are of ambiguous origin (page 784).

Kramer (1983)
Theory of competence: voters respond to the policy component but not the exogenous component that is beyond governmental control.

Alesina and Rosenthal (1995)
Expectations and uncertainty: the study develops a competence model and relates the theoretical concepts of *expectations* about inflation and *uncertainty* about incumbent competence to the empirical analogues of *conditional expectations* and *error-in-variables* regression.

Suzuki and Chappell (1996)
Uncertainty: the study develops different measures for uncertainty in which voters are able to differentiate between permanent and temporary changes in economic growth in evaluating incumbent competence.

Mebane (2000)
Expectations: the study models coordination in elections in the form of a noncooperative rational expectations equilibrium among voters. Policy moderation is achieved given voters' private knowledge about the election, rather than common knowledge alone.

Duch and Stevenson (2008)
Uncertainty and signal extraction: the study applies different measures for uncertainty that distinguish electorally dependent decision makers from non-electorally dependent ones. The analogues are then applied to a variety of error-in-variables (signal extraction) problems with respect to the attribution of responsibility.

Duch and Stevenson (2010)
Uncertainty and signal extraction: the study applies the error-in-variables analogue to signal extraction problems in the contexts of uncertainty over domestic and global economic volatilities, and openness to global trade. The study provides a competence signal explanation for the variations in the magnitude of economic vote.

Achen (2012)
Uncertainty and error-in-variables: the study relates the behavioral concept of uncertainty to an error-in-variables problem. It differs from the basic framework by relating growth rate, instead of level of growth, to competence. In addition, instead of the conventional MA(1) structure for competency, the new model implies different lag structures to generate more accurate estimates.

Alcañiz and Hellwig (2011)
Uncertainty and signal extraction: the empirical study makes contributions by demonstrating how signaling models can be used to better understand responsibility attribution by voters on the one hand, and the exoneration of blame for the political elites on the other.

Fig. 1. Economic voting cumulation: A sample.

Based on survey data in which respondents are allowed to select from a large set of response options with regard to responsibility attribution, Alcañiz and Hellwig make contributions by demonstrating how signaling models can be used to better understand the attribution of responsibility. On the one hand, they show that voters in developing economies seek to reduce uncertainty with regard to responsibility attributions based on their knowledge about their countries' positions in the global economy (Alcañiz and Hellwig, 2011: 390). On the other hand, the authors point out that political elites are more likely to avoid blame in order to hold on to power (Alcañiz and Hellwig, 2011: 408).

A more recent study by Achen (2012) develops a new model to better understand the dynamics of voter retrospection. Following Alesina and Rosenthal (1995), Achen models growth as a measurement error problem with the voter having uncertainty about how much of the economic performance is due to incumbent competence. However, instead of modeling incumbent competence as a moving average of an (MA(1)) process, Achen *extends voter evaluation of incumbent competence to encompass the full term of an administration.* He assumes the more volatile the economy, the more persistent voter memory is with respect to evaluating competence.

The choice of dependent variable is also different in Achen's model: voters relate competence with growth rates, not income levels (Achen, 2012: 9). This chance can be found in Equation (7). The MA(1) process similar to the growth model in the Alesina and Rosenthal model, except now the dependent variable is g_t^*, the unobserved true growth rate:

$$g_t^* = g_{t-1}^* + \delta_t + \rho\delta_{t-1} \quad 0 < \rho < 1, \tag{7}$$

where:

g_t^* denotes the unobserved true growth rate in period *t*,
g_{t-1}^* denotes the observed true growth rate in period *t−1*.

The competence parameters, δ_t and $\delta_{t-1,}$ are similar to the parameters μ_t and μ_{t-1} in Equation (3). The voter is assumed to know ρ but not δ_t.

Let the voter have the ability to observe a "noisy" version of the true growth rate g_t^* in period *t* containing measurement error:

$$g_t = g_t^* + e_t. \tag{8}$$

In (8), g_t, is the observed growth rate in period *t*; e_t is the error in the voter's perception of growth in period *t*, with mean zero and variance (σ_e^2). The term e_t is uncorrelated with e_{t-1} and all of its other past values. Substituting (g_{t-1} - e_{t-1}) forg_{t-1}^* in (7) and then rewriting (8) gives:

$$g_t = g_{t-1} + \delta_t + \rho\delta_{t-1} + e_t - e_{t-1}, \tag{9}$$

Equation (9) can be simplified to:

$$g_t = g_{t-1} + w_t, \tag{10}$$

where: $w_t = \delta_t + \rho\delta_{t-1} + e_t - e_{t-1}$.

Since w_t is the sum of two MA(1) processes, it can be rewritten as an MA(1) process itself:

$$w_t = \varepsilon_t - \gamma\varepsilon_{t-1}, \tag{11}$$

where: ε_t has zero mean and variance (σ_ε^2).

Equation (11) can be expressed as:

$$\varepsilon_t - \gamma\varepsilon_{t-1} = \delta_t + \rho\delta_{t-1} + e_t - e_{t-1}. \tag{12}$$

Achen (2012: 6–7) then solves for γ, the geometric weight parameter, in Equation (12) and relates it to the voter's forecast:

$$g_t - g_{t-1} = \varepsilon_t - \gamma\varepsilon_{t-1}. \tag{13}$$

Using lag operators there exists a white noise process (13) can be written as:

$$g_t - g_{t-1} = (1 - \gamma L)\varepsilon_t. \tag{14}$$

It follows that:

$$\varepsilon_t = \frac{g_t}{(1-\gamma L)} - \frac{g_{t-1}}{(1-\gamma L)}. \tag{15}$$

Multiplying out gives:

$$\varepsilon_t = (g_t + \gamma g_{t-1} + \gamma^2 g_{t-2} + \ldots) - (g_{t-1} + \gamma g_{t-2} + \gamma^2 g_{t-3} + \ldots). \tag{16}$$

Arranging terms, taking expectations, and truncating lags greater than 4 gives the best forecast for economic growth as follows:

$$\widehat{g}_t = (1-\gamma)\sum_{k=1}^{4}\gamma^{k-1}g_{t-k}, \tag{17}$$

where:

$\widehat{g}_t$ is the expected GDP growth from period *t−1* to period *t*, formed at the end of period *t−1*, γ is the geometric weight parameter.

The growth model in expression (17) contrasts with Alesina and Rosehnthal. Voters in Achen's model determine incumbent competence in three ways (Achen, 2012: 10):

1. The persistent case ($\gamma \to 1$): geometric weights decline slowly, implying that nearly an equal weighting of all years in evaluating the growth forecast.
2. The myopic case ($\gamma \to 0$ and positive): the weights decline rapidly, so that only the most recent period (year) matters.
3. The "in between" case ($0 < \gamma < 1$): the weights decline gradually in voter retrospection.

Achen conducts a preliminary test of his model using data from the highly volatile economic circumstances of Montana wheat-growing counties in the 1930s. The results show voter memory in evaluating President Hoover's competence persists. The author contends this contributed (at least partly) to Hoover's defeat in his 1932 even though the economy improved during the latter part of his presidency.

Achen's model provides for more accurate estimates through the choices of the lag parameter in the specifications (i.e., pure myopia, persistence, or something in between). The study also points to future directions in the cumulation process: for example, if competence is modeled as an AR(1) process, the model will result in an ARMA(1,2) (autoregressive-moving-average) lag structure. Other ARMA structures can also be used for empirical testing by allowing the four lag coefficients to take on different values (Achen, 2012: 17).

Yet another refinement to Alesina and Rosenthal's can be found in the work of Walter Mebane (2000). He provides a more sophisticated theory of policy moderation by voters than the basic framework of Alesina and Rosenthal. According to the author, his model is a useful way in generalizing the approach used by Alesina and Rosenthal (Mebane, 2000: 40), by depicting the important difference in the information that voters hold. In Alesina and Rosenthal's model, voters possess common knowledge about the elections and behave strategically to achieve policy moderation but they do not have any private information about election outcomes (Mebane, 2000: 38). The individual voters in Mebane's model, on the other hand, *possess both common knowledge and private information about the elections – such as individual policy preferences and perceptions of the candidates*. It follows that the larger the discrepancy between the voter's policy ideal point and the expected policy given the election outcome, the bigger the expected loss for the voter.

While the basic framework of Alesina and Rosenthal treats the expected policies and expected election outcomes by all voters to be identical, the new model explains much more about the information that voters have, by allowing *the expected policies and the expected election outcomes to vary across voters* (Mebane, 2000: 39–40). Based on his model, Mebane further posits that policy moderation is achieved through coordination in the form of a noncooperative rational expectations equilibrium among voters. Empirical evidence obtained from the model's analysis sheds new light by showing that there is a small but significant proportion of voters who vote a split ticket in order to improve the chances of policy moderation.

8. Conclusion

The EITM framework offers a linkage between formal modeling and empirical analysis. Some current methodological practices inhibit the cumulation of knowledge due, in part to the ongoing disconnect between formal and empirical modelers. The status quo is one where isolation of fields and sub-fields is dominant. Such compartmentalization exacerbates the separation between theoretical and empirical models impairing the potential for scientific advancement.

We believe significant scientific progress can be made by unifying formal and empirical modeling. This methodological unification also leads to the use of an ever increasing set of behavioral concepts. Applying the EITM framework means new and better ways will be discovered to model human behavior. The repeated application of competing analogues raises the possibility of new thinking how humans act, but now with a sense there is a rigor in putting these new behavioral developments to the test.

It is also important to avoid the trap of conducting current debates using our past and current training as the basis for the debate. Straight jacketed thinking – along the lines of Clarke and Primo – translates to an avoidance in dealing with known weaknesses in our current practices. Instead, what is needed is the belief that unifying modeling practices and tools can be pushed further and these new ideas survive – for a limited time – if they improve upon past and current practice.

To work these new innovations are certain to possess properties we know to enhance understanding, whether it involves measurement, better ways to characterize human behavior, sampling, and more. But bear in mind the new ways of analyzing important and numerous social science research questions must also be designed to preserve and enhance the dialogue between the inner workings of a system and tests.

References

Achen, C.H., 2002. Toward a new political methodology: microfoundations and ART. Annu. Rev. Politi. Sci. 5 (1), 423–450.
Achen, C.H., 2012. When Is Myopic Retrospection Rational?. Accessed 01.22.14, from. http://www.class.uh.edu/hcpp/EITM/Achen.pdf.
Ahmed, F.Z., 2012. The perils of unearned foreign income: aid, remittances, and government survival. Am. Politi. Sci. Rev. 106 (1), 146–165.
Akerlof, G.A., 2002. Behavioral macroeconomics and macroeconomic behavior. Am. Econ. Rev. 92 (3), 411–433.
Alcañiz, I., Hellwig, T., 2011. Who's to blame? The distribution of responsibility in developing democracies. Br. J. Politi. Sci. 41 (2), 389–411.
Aldrich, J.H., 1993. Rational choice and turnout. Am. J. Politi. Sci. 37 (1), 246–278.
Aldrich, J.H., Alt, J.E., Lupia, A., 2008. The eitm approach: origins and interpretations. In: Box-Steffensmeier, J.M., Brady, H.E., Collier, D. (Eds.), The Oxford Handbook of Political Methodology. Oxford University Press, New York, pp. 828–843.
Alesina, A., Rosenthal, H., 1995. Partisan Politics, Divided Government, and the Economy. Cambridge University Press, New York.
Arena, P., 2015. Crisis bargaining, domestic opposition, and tragic wars. J. Theor. Polit. 27 (1), 108–131.
Arrow, K.J., 1963. Social Choice and Individual Values, second ed. John Wiley & Sons, Inc., New York.
Bapat, N.A., 2005. Insurgency and the opening of peace process. J. Peace Res. 42 (6), 699–717.
Blonigen, B., Wang, M., 2005. Inappropriate pooling of wealthy and poor countries in empirical FDI studies. In: Moran, T.H., Graham, E.M., Blomstrom, M. (Eds.), Does Foreign Direct Investment Promote Development? Institute for International Economics, Washington, pp. 221–244.
Bonica, A., 2013. Ideology and interests in the political marketplace. Am. J. Politi. Sci. 57 (2), 294–311.
Bowles, S., 2008. Policies designed for self-interested citizens may undermine “the moral sentiments”: evidence from economic experiments. Science 320 (5883), 1605–1609.
Brians, C.L., Grofman, B., 1999. When registration barriers fall, who votes? An empirical test of a rational choice model. Public Choice 99 (1–2), 161–176.
Camerer, C.F., Fehr, E., 2006. When does “economic man” dominate social behavior? Science 311 (5757), 47–52.
Centola, D., Willer, R., Macy, M., 2005. The emperor's dilemma: a computational model of self-enforcing norms. Am. J. Sociol. 110 (4), 1009–1040.
Clarke, K.A., Primo, D.M., 2007. Modernizing political science: a model-based approach. Perspect. Polit. 5 (4), 741–753.
Clarke, K.A., Primo, D.M., 2012. A Model Discipline: Political Science and the Logic of Representations. Oxford University Press, New York.
Converse, P.E., 1964. The nature of belief systems in mass publics. In: Apter, D.E. (Ed.), Ideology and Discontent. Free Press, New York, pp. 206–261.
Converse, P.E., 1969. Of time and partisan stability. Comp. Polit. Stud. 2 (2), 139–171.
Cox, J., Friedman, D., Gjerstad, S., 2007. A tractable model of reciprocity and fairness. Games Econ. Behav. 59 (1), 17–45.
Crawford, S.E.S., Ostrom, E., 1995. A grammar of institutions. American Political Science Review 89, 582-600. Revision, 2005. In: Ostrom, E. (Ed.), Understanding Institutional Diversity. Princeton University Press, Princeton, NJ, pp. 137–174.
Downs, A., 1957. An Economic Theory of Democracy. Harper and Row, New York.
Duch, R.M., Stevenson, R., 2008. Voting in Context: How Political and Economic Institutions Condition Election Results. Cambridge University Press, Cambridge.
Duch, R.M., Stevenson, R., 2010. The global economy, competency, and the economic vote. J. Polit. 72 (1), 105–123.
Fehr, E., Fischbacher, U., Kosfeld, M., 2005. Neuroeconomic foundations of trust and social preferences: initial evidence. Am. Econ. Rev. 95 (2), 346–351.
Fiorina, M.P., 1978. Economic retrospective voting in American national elections: a micro-analysis. Am. J. Politi. Sci. 22 (2), 426–443.
Friedman, M., 1957. A Theory of the Consumption Function. Princeton University Press, Princeton.
Granato, J., 2005. Scientific progress in quantitative political economy. Politi. Econ. 12 (4), 1, 3, 9, 11–13, 20.
Granato, J., Lo, M., Wong, M.C.S., 2010a. A framework for unifying formal and empirical analysis. Am. J. Politi. Sci. 54 (3), 783–797.
Granato, J., Lo, M., Wong, M.C.S., 2010b. The empirical implications of theoretical models (EITM): a framework for methodological unification. Política Gob. XVII (1), 25–57. The English translation can be found here. http://www.class.uh.edu/hcpp/jimgranato/GranatoLoWongPoliticayGobierno.pdf.

Granato, J., Lo, M., Wong, M.C.S., 2011. Modeling and testing the diffusion of expectations: an EITM approach. Elect. Stud. 30 (3), 389–398.
Granato, J., Scioli, F., 2004. Puzzles, proverbs, and omega matrices: the scientific and social consequences of the empirical implications of theoretical models (EITM). Perspect. Polit. 2 (2), 313–323.
Keynes, J.M., 1939. Professor Tinbergen's method. Econ. J. 49 (195), 558–577.
Kiewiet, D.R., Rivers, D., 1984. A retrospective on retrospective voting. Polit. Behav. 6 (4), 369–393.
Kinder, D.R., Kiewiet, D.R., 1979. Economic discontent and political behavior: the role of personal grievances and collective economic judgments in congressional voting. Am. J. Politi. Sci. 23 (3), 495–527.
Kinder, D.R., Kiewiet, D.R., 1981. Sociotropic politics: the American case. Br. J. Politi. Sci. 11 (2), 129–161.
Kramer, G., 1983. The ecological fallacy revisited: aggregate- versus individual-level findings on economics and elections, and sociotropic voting. Am. Politi. Sci. Rev. 77 (1), 92–111.
Kyburg, H.E., 1988. The justification of deduction in science. In: Grunbaum, A., Salmon, W.C. (Eds.), The Limitations of Deduction. Pittsburgh Series in Philosophy and History. University of California Press, Berkeley, pp. 61–94.
Leamer, E.E., 1983. Let's take the con out of econometrics. Am. Econ. Rev. 73 (1), 31–43.
Leamer, E.E., 2010. Tantalus on the road to asymptopia. J. Econ. Perspect. 24 (2), 31–46.
Lopez, M.C., Murphy, J.J., Spraggon, J.M., Stranlund, J.K., 2009. Comparing the Effectiveness of Regulation and Pro-social Emotions to Enhance Cooperation: Experimental Evidence from Fishing Communities in Colombia. University of Massachusetts Amherst Department of Resource Economics Working Paper No. 2009-5. Accessed 02.01.14, from SSRN. http://dx.doi.org/10.2139/ssrn.1476848.
Lucas Jr., R.E., 1973. Some international evidence on output-inflation tradeoffs. Am. Econ. Rev. 63 (3), 326–334.
Lucas Jr., R.E., 1976. Econometric policy evaluation: a critique. Carnegie-Rochester Conf. Ser. Public Policy 1, 19–46.
Lucas Jr., R.E., 1988. On the mechanics of economic development. J. Monetary Econ. 22, 3–42.
McCabe, K., Smith, V., 2001. Goodwill accounting in economic exchange. In: Gigerenzer, G., Selten, R. (Eds.), Bounded Rationality: the Adaptive Tool Box. MIT Press, Cambridge, pp. 319–342.
Mebane Jr., W.R., 2000. Coordination, moderation, and institutional balancing in American presidential and House elections. Am. Politi. Sci. Rev. 94 (1), 37–57.
Minford, P., 1992. Rational Expectations Macroeconomics: an Introductory Handbook, second ed. Wiley-Blackwell, West Sussex.
Olson Jr., M., 1965. The Logic of Collective Action. Harvard University Press, Cambridge.
Ostrom, E., 2009. A general framework for analyzing sustainability of social-ecological systems. Science 325 (5939), 419–422.
Ostrom, E., 2010. Revising theory in light of experimental findings. J. Econ. Behav. Organ. 73 (1), 68–72.
Ostrom, E., Gardner, R., Walker, J., 1994. Rules, Games, and Common-pool Resources. University of Michigan Press, Ann Arbor.
Ostrom, E., Walker, J., 1991. Communication in a commons: cooperation without external enforcement. In: Palfrey, T.R. (Ed.), Laboratory Research in Political Economy. University of Michigan Press, Ann Arbor, pp. 287–322.
Ostrom, E., Walker, J., Gardner, R., 1992. Covenants with and without a sword: self-governance is possible. Am. Politi. Sci. Rev. 86 (2), 404–417.
Palfrey, T.R., Rosenthal, H., 1985. Voter participation and strategic uncertainty. Am. Politi. Sci. Rev. 79 (1), 62–78.
Pfleiderer, P., 2014. Chameleons: the Misuse of Theoretical Models in Finance and Economics. Accessed 10.02.14, from. https://www.gsb.stanford.edu/sites/default/files/research/documents/Chameleons%20-The%20Misuse%20of%20Theoretical%20Models%20032614.pdf.
Penn, E.M., 2008. Citizenship versus ethnicity: the role of institutions in shaping identity choice. J. Polit. 70 (4), 956–973.
Poteete, A.R., Janssen, M.A., Ostrom, E., 2010. Working Together: Collective Action, the Commons, and Multiple Methods in Practice. Princeton University Press, Princeton.
Riker, W.H., 1958. The paradox of voting and congressional rules for voting on amendments. Am. Politi. Sci. Rev. 52 (2), 349–366.
Sally, D., 1995. Conversation and cooperation in social dilemmas: a meta-analysis of experiments from 1958 to 1992. Ration. Soc. 7 (1), 58–92.
Salmon, W.C., 1988. Introduction. In: Grünbaum, A., Salmon, W.C. (Eds.), The Limitations of Deductivism. University of California Press, Berkeley, pp. 1–18.
Sargent, T., 1976. The observational equivalence of natural and unnatural rate theories of macroeconomics. J. Politi. Econ. 84 (3), 631–641.
Schmitt, P., Swope, K., Walker, J., 2000. Collective action with incomplete commitment: experimental evidence. South. Econ. J. 66 (4), 829–854.
Sims, C.A., 1980. Macroeconomics and reality. Econometrica 48 (1), 1–48.
Solow, R.M., 1956. A contribution to the theory of economic growth. Q. J. Econ. 70 (1), 65–94.
Stokes, D., 1963. Spatial model of party competition. Am. Politi. Sci. Rev. 57 (2), 368–377.
Svolik, M.W., 2009. Power sharing and leadership dynamics in authoritarian regimes. Am. J. Politi. Sci. 53 (2), 477–494.
Suzuki, M., Chappell, H.W., 1996. The rationality of economic voting revisited. J. Polit. 58 (1), 224–236.
Wagner, R.H., 2001. Who's Afraid of "Rational Choice Theory"?. Accessed 12.22.13, from. http://hw.webhost.utexas.edu/papers/rct.pdf.
Zellner, A., 1984. Basic Issues in Econometrics. University of Chicago Press, Chicago.