CrossMark

Comparative Politics and Area Specialization: The Hidden Costs of Analytic Transparency

Sean Yom[1]

Abstract New transparency initiatives emerging from Western political science make an ill-fitting standard for global comparativists who serve as experts in countries and regions—that is, area specialists. This study explores the origins of the current push for greater research transparency, including its underlying norms and genesis in the natural sciences. Noting its strict adherence to a model of deductive proceduralism, I then show how area-focused comparative political scholarship clashes with newfound demands for analytic transparency due to fundamental differences in its real-world practice—that is, how explanations are constructed and evidence is analyzed. Rather than a linear step-by-step approach emphasizing hypothesis confirmation, much area studies work thrives through iterative engagements between theory and evidence in which researchers craft persuasive explanations by repeatedly revising propositions, reconsidering data, and questioning assumptions over time. Squaring this reality with the idealization of enhanced transparency means one of two things: the diminishment of area studies within comparative politics, or the collective lying of many scholars when it comes to divulging how they came to their ultimate causal conclusions.

1 Introduction

Over the past several years, the subfield of comparative politics has given rise to new debates over imposing greater standards of *transparency*. These discussions originated within the discipline's English-speaking Western core and have

✉ Sean Yom
seanyom@temple.edu

1   Department of Political Science, Temple University, Philadelphia, PA 19122, USA

 Springer

magnanimous implications for future practice and education. From annual meetings of the American Political Science Association (APSA) and smaller workshops, to various online statements and message boards, a growing number of scholars have wondered how political science ought to revamp its standards of data access and research transparency (DA–RT) to encourage more openness and reliability in scholarship. Though discussions over DA–RT began in biology, medicine, and the physical sciences, it has penetrated political science to a degree not witnessed in neighboring fields like sociology and history.

Some political scientists had long urged their colleagues to be more transparent, for instance, making their data readily available and precisely delineating the value of each scholarly reference (McDermott 2010; Moravcsik 2010). However, the massive translation of this norm into new rules and procedures only began in 2012 with changes made to APSA's *Guide to Professional Ethics in Political Science*, followed by the 2014 *Journal Editors Transparency Statement* (JETS), which committed major signatory journals to develop new author instructions and publication policies in accordance with DA–RT guidelines.[1] The July 2015 Transparency and Openness Promotion statement released by the Center for Open Science added parallel momentum to the DA–RT initiative within political science, highlighting the need for greater disclosure and openness in all aspects of scholarship—i.e., citation standards, data sharing, coding procedures, research design, hypothesis generation, analysis plans, and replication.[2] Debates about the feasibility and ethicality of DA–RT measures have invoked symposia in major publications, such as the spring 2014 issue of *PS: Political Science and Politics*, the spring 2015 issue of the *APSA Section for Qualitative and Multi-Method Research (QMRR)* Newsletter, and the April 2016 issue of the *APSA Section for Comparative Politics Newsletter*. Online forums have discussed these issues, among them *Dialog on DA–RT*,[3] *Qualitative Transparency Deliberations*,[4] and *PSNow*.[5]

These dialogs are thriving, and sometimes rancorous. Many English-speaking senior political scientists, mostly based in the US, have advocated the principles of DA–RT to guide future research and publication practices at prominent journals.[6] An eclectic front of critics, including area studies comparativists, has fired back. Many note the logistical difficulties and ethical implications of imposing uniform standards in terms of the first two elements of DA–RT—*data access* and *production transparency*. For instance, apropos data access, some find problematic the requirement that all data used in published work to be stored in publicly available trusted repositories, given that data can come in forms as disparate as numeric

---

[1] These statements can be read in entirety on the DA–RT Statement website: http://www.dartstatement. org.

[2] See further the Transparency and Openness Guidelines online: http://cos.io/top.

[3] See http://dialogueondart.org.

[4] See http://www.qualtd.net.

[5] See http://www.politicalsciencenow.com/category/da-rt.

[6] For instance, in spring 2015, the *American Journal of Political Science* (a JETS signatory) mandated a new submission requirement: all results accepted for publication needed to be independently replicated and verified utilizing the data and other materials submitted by the author.

rectangular sets, sensitive historical archives, and field-based interviews conducted under confidentiality. Regarding production transparency, others wonder how explicit calls to clarify all procedures used to collect data will affect researchers utilizing different methodologies, ranging from computational statistics and formal modeling to qualitative techniques like ethnographic interpretation, historical comparison, and case-based process-tracing.

There is a third dimension of transparency, though one that incurs profound practical and normative conflict—*analytic transparency*. As the revised APSA Ethics Guide states, political scientists making "evidence-based knowledge claims" now have the obligation to provide "a full account of how they draw their analytic conclusions from the data, i.e., clearly explicate the links connecting data to conclusions." This study intervenes at this inflection point. It questions the viability of DA–RT's call for analytic transparency from the ambit of *area specialization*— that is, from the viewpoint of comparativists who specialize in specific countries and regions, whose work primarily deals with a few number of cases, and whose careers require leveraging (and constantly demonstrating) superior expertise in these places. It shows that the prevailing disciplinary understanding of analytic transparency signals severe misunderstanding of how area-specific comparative research actually operates in the real world.

This study proceeds in three parts: first, DA–RT's origins and the impulse for scientific legitimacy; second, the nature of area specialization, with its mandate for iteration and clutter that runs against the idealized image of deductive proceduralism; and third, the manifest differences in hypothesis-testing, the core element of scientific understanding, within area studies and other fields.

2 The Rise of DA–RT

Political scientists cannot be trusted to conduct research in honest ways, and peer-reviewers at journals cannot be trusted to catch all potential improprieties. These, at heart, are the two presumptions of DA–RT. *Ipso facto*, absent new standards encouraging greater transparency, existing political science scholarship will continue to be riddled with errors and misconduct that muck up the scientific method.

It is useful, given this starting point, to trace backward how the impetus for open data access and greater transparency began outside of political science. Especially in statistics and psychology, scholars have always questioned the rigor and honesty of their field's practices with every generation (e.g., Sterling 1959). Yet, in the twenty-first century, a new wave of doubt within the biological and physical sciences descended upon whether scholars were truly undertaking their research in objective ways. John Ioannidis' meta-reviews suggesting that most published findings and causal associations in medicine were false represent the most prominent example (Ioannidis 2005, 2008). His underlying claim was implicit and damning: the prejudices of researchers often interfered with the dispassionate unfolding of the scientific method, sometimes knowingly and sometimes unconsciously, due to the natural impulse to generate positive results and "prove" an initial hypothesis

(Simmons et al. 2011). More specifically, some analysts in psychology (Masicamp and Lalande 2012) and political science (Gerber and Mahotra 2008) found evidence of "publication bias" in studies utilizing statistical methods, which suggests the tampering of quantitative data or model-fitting in order to squeeze out significant results that would have otherwise not been obtained. Aggregative meta-studies have thrown more fuel upon the fire, showing that projects with "negative" findings (that is, where the original hypothesis or argument was not corroborated by the data) are disappearing from the ranks of journal publications in most scientific fields (Fanelli 2012). Either scholars were simply right most of the time, or else the strategic pressures of publishing combined with various external biases have conspired to encourage investigators to massage out positive findings while quietly discarding negative ones, even if they are closer to reality.

Perhaps, the best way to ascertain the rigor of research practices is attempting to reproduce published results. After all, if scientists followed the steps outlined in finished studies with the same data and methods, then surely, they would generate similar findings most of the time. Yet, this is not the case. The Reproducibility Project, a collaborative endeavor involving nearly 300 scholars attempting to replicate 100 published experimental and observational psychological studies, found only a little over a third of the replications led to statistically significant results (Open Science Collaboration 2015: 943). Even before this well-publicized result, numerous other medical teams had similarly failed to reproduce results from published clinical trials and experiments (Prin et al. 2011; Pusztai et al. 2013). Anecdotally, the same problems can be found in the empirical subfields of political science. As Gary King and others have surmised (and as many others privately whispered), it was incredibly difficult to replicate many results from empirical studies published in leading political science journals, whether they were quantitative or qualitative in methodology (King 1995). This "replication crisis" is not based on the law of averages, but instead reflects the modal outcome. The occasional public scandal involving political scientists fabricating data certainly has not helped matters. [See, for instance, *Science* journal's retraction of LaCour and Green 2014)]. In fairness, though fraudulent research behavior pollutes a wide variety of scientific fields. A classic example is *Lancet's* 2010 retraction of the infamous 1998 study linking MMR vaccination to autism, which returned incredible evidence of data-mining and selective reporting (Deer 2011). Indeed, one needs only peruse *Retraction Watch* website to witness a veritable "hall of shame" of scholarship discredited due to misconduct.[7]

In this context where leading empirical political scientists worried about the prevalence of improprieties in their field—"star-gazing," "p-fishing," curve-fitting, cherry-picking data, altering hypotheses, massaging results, orchestrated coding, reinventing studies to make it seem as if they "predicted" the obtained results all along—DA–RT's rise is understandable. It negates the gloomy fear that has settled above the English-speaking comparative political science scene: that we are generating knowledge, creating theories, and advancing careers on a foundation of questionable findings that make a mockery of scientific ideals. In this milieu, DA–

---

[7] See http://www.retractionwatch.com.

RT is but one entry in an alphabetic soup of new programs and networks dedicated to enhancing transparency across the social sciences by sharing data, preregistering hypotheses, and encouraging replicability—among them, the COS, OpenAIRE, EGAP, BITSS, Project TIER, Dataverse Network, and ReplicationWiki.

Critically, some DA–RT measures (such as changes in the APSA Ethics Guide and the JETS) within political science resemble steps taken within the biological and physical sciences. For instance, data depositories have popped up in many disciplines. One impressive example is the NSF-funded Paleobiology Database, which globally covers all collection-based occurrence and taxonomic data from any epoch.[8] Likewise, the *Journal of the American Medical Association* (JAMA) recently outlined new requirements from the International Committee of Medical Journal Editors (ICMJE) that would mandate more data-sharing, including metadata, for interventional clinical trials (JAMA 2016). Since the ICMJE already required the preregistration of all clinical trials prior to enrollment of the first participant, this measure was designed to enhance reproducibility, prevent selective reporting, and reduce publication bias. As another example, in October 2014, the Coalition for Publishing Data in the Earth and Space Sciences (COPDESS) began providing an organizational framework to help share data through common facilities and jointly implement transparency procedures across constituent field like astronomy and ecology.

Yet, it is within the policies of peer-reviewed journals, the gatekeepers of scientific research, where the most visible changes have occurred. In January 2014, *Science* announced that for preclinical studies, the journal would be adopting recommendations from the US National Institute for Neurological Disorders and Stroke (McNutt 2014). Thus, authors would need to indicate whether a pre-experimental plan for data analysis, including contingencies for outliers; whether they conducted sample size estimation to ascertain the signal-to-noise ratio; and validation that the experimenter was blind to the conduct of the experiment. In April 2013, *Nature* abolished space restrictions on the methodology sections of accepted articles. Authors also needed to successfully complete a reporting checklist that required disclosure of all research steps and critical parameters—e.g., cell line identities, antibody reagent numbers, human and animal protocols, sampling and blinding techniques, and divulging of all pre-established hypotheses and criteria, especially for the experimental work (*Nature* 2013: 398).

Many others have followed suit. In September 2014, *Developmental Cell* introduced new supplemental guidelines that called for authors to upload a detailed account of their research protocols in addition to their primary data, which would complement existing editorial measures to filter out improperly conducted studies, such as screening figure images for inconsistencies (Wainstock 2014). Some journals have even tried tinkering with the peer-review process. For example, in 2013, *GigaScience* began allowing for open peer-review without any blinding of referees (Edmunds 2013: 1). Likewise, in December 2015, *Nature Communications* began allowing authors to publish all peer-review documentation, including reviewer comments and rebuttal letters.

---

[8] See http://paleobiodb.org.

Hundreds of other journals, forums, and associations are changing their practices in conformity with this new wave. I present this selection here to contextualize debates over DA–RT in political science. DA–RT takes place against the backdrop of consensus within the biological and physical sciences regarding the need for greater transparency in research. That consensus holds that all scholars need to be forthcoming in disclosing their samples, methods, reagents, theories, codes, protocols, and techniques. In this sense, proponents of DA–RT enjoy an ethically incontestable stance. Transparency is a public good, and science belongs in the public domain. It makes sense, then, that the two should be married in a meaningful way. As several APSA presidents declared online, "Our legitimacy as scholars and political scientists in speaking to the public on issues of public concern rests in part on whether we adopt and maintain common standards for evaluating evidence-based knowledge claims" (Hochschild et al. 2015).

However, what remains to be seen is whether the drive for scientific authenticity and disciplinary legitimacy will endanger certain kinds of scholarship in the future, such as area specialists for whom analytic transparency connotes very different requirements than for other empiricists. The source of this misfit is that whereas DA–RT and similar initiatives assume that scholars will (or at least should) implement deductive step-by-step procedures over the course of research, in reality many scholars within comparative politics are far more eclectic and iterative in how they construct causal explanations. Area specialists pursue causal inferences, and wish to enrich (or create anew) theories that help explain why things occur in distant lands. However, the process of exercising and working with knowledge over other regions, states, and societies require research practices that extend far beyond the model of practices envisaged by DA–RT.

3 Iterative Components of Area Specialization

There is no need to regurgitate the epistemological battles of the 1990s and early 2000s, when advocates of area studies and exponents of generalized theory repeatedly clashed over the direction of comparative politics (Bates 1997; Hanson 2009). Today, the discipline has matured enough—and, worldwide, grown large enough—to encompass virtually any methodological or theoretical orientation. By area studies, I mean the collective effort to "document the existence, internal logic, and theoretical implications" of societies, state, or regions found beyond one's own cultural experience (Szanton 2004). Whether by that classic moniker or trendier names like "area-focused scholarship" and "global and regional studies," scholarly specialization of foreign places has long had a cherished place in comparative politics (Sil 2009). Indeed, it still cognitively dominates how many Western political scientists treat the subfield. For instance, many US-based departments still hire new comparativists-based region of expertise, ensuring that there are no overlaps with existing geographic specializations within the faculty. Career stability can thus depend upon maintaining superior expertise and knowledge accumulated from that at least one major geographic zone, or specific cases of countries and societies (or even smaller units) within that regional space.

Area specialization, however, is not without costs. Among the resources most consumed is time—time to learn foreign languages and collect data in the field, the latter of which in turn requires cultivating trust and friendship with those already living there. The handmaiden to securing such deep expertise is familiarity, a conditionalized level of personal comfort in dealing with another culture or society, different political or social values, and a novel historical vocabulary. Familiarity comes after repeated exposure to the unfamiliar. On balance, a Ph.D. candidate who makes repeated trips to Ecuador over 2 years totaling 9 months will be generally more familiar with Ecuadorian state and society than a peer who may have collected an equally impressive array of data but resided there for only 3 weeks. Across the subfield, the leading experts in various regions and topics—Middle East authoritarianism, East Asian development, sub-Saharan African conflict, Central American state-building, South American public goods provision, South Asian industrialization, Southeast Asian democratization—are those who maintain contact with the ever-changing conditions of their geographic targets.

Another hallmark of area specialization within comparative politics is the productive value of *iteration*. By iteration, I mean the flexibility of moving between theory and data, homeland and the field, in a constant effort to recalibrate one's knowledge and reconcile theoretical arguments with evidence on the ground. This requires willingness to violate the rules of austere hypothesis-testing which in past work I called the "deductive template" (Yom 2015a: 620). The deductive template shares familial attributes with the philosophical model of hypothetico-deductive investigation, which observers like Kevin Clarke and David Primo see as the "gold standard" in empirical political science despite its origins in nineteenth century Newtonian physics (Clarke and Primo 2012: 26). David Waldner portrays this deductive worldview as the procedural manifestation of the Popperian quest for hypothesis confirmation, one that sees the goal of all serious research as the rigorous testing of causal propositions (Waldner 2007). In any guise, the deductive template connotes the image of political scientists working in laboratory-like conditions, patiently deducing a falsifiable hypothesis from existing theories as well as the predictive implications from the hypothesis, continuing with collecting data and cases, culminating in the comparison of actual data with the hypothetical predictions, and ending with an austere conclusion that the hypothesis was either disconfirmed or confirmed (Yom 2015a: 622–623). Figure 1 outlines this.

The problem with this deductive template lies with its inherent proceduralism. Most generally, pressures to adhere to this linear, step-by-step logic of the deductive template can be injurious to theory-building, because it construes all propositions as either true or false, right or wrong. Yet, this ignores that many representations of reality utilized within political analysis, from complex game-theoretic models to extended narrative-driven case studies, can still hold useful information (including descriptive data) about actors, choices, and causation that would otherwise be ignored in the ritualistic hunt to corroborate or discredit hypotheses. Second, because this paradigm (and the journals, foundations, and conferences imbibing it) prizes the goal of successfully confirming innovative hypotheses, it strongly discourages occurrences of negative or null results, even though these can hold
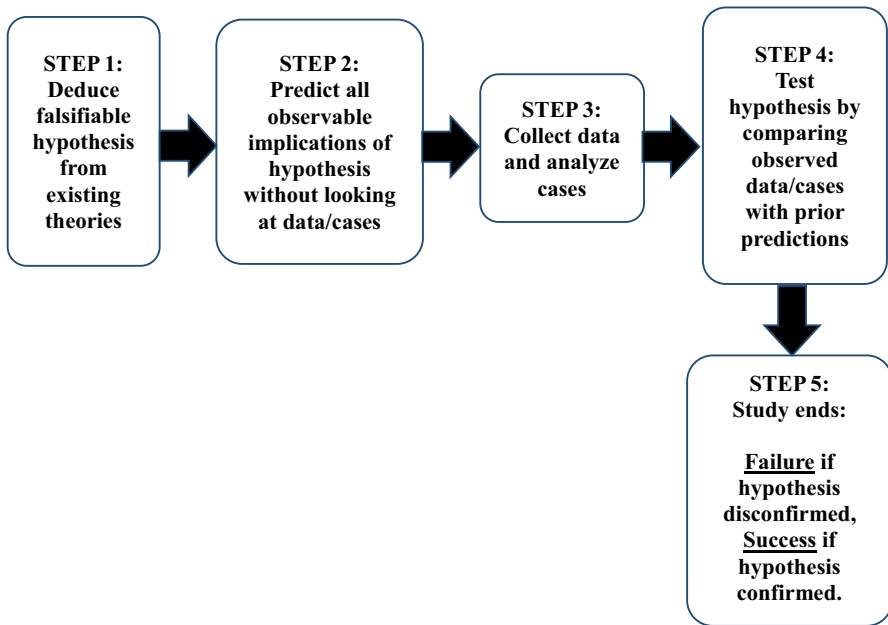
Fig. 1 The deductive model of research

decisive information about causal effects. Herein, lays the "desk drawer" problem (Rosenthal 1979).

Most of all, the deductive template can broach on the realm of fantasy to area specialists working through complex projects, appearing like a neo-Kuhnian wonderland of "normal science" populated by merry Popperians falsifying their way through life. If this is science, it is science fiction—a fiction not only hoisted upon graduate students but also a laughably dishonest representation of what so many of us, as scholars, really do when completing dissertations, securing grants, writing articles and books, and presenting at conferences. As David Laitin has noted, "Nearly all research in political science… goes back and forth between a set of theories and an organically growing data archive with befuddled researchers trying to make sense of the variance on their dependent variables" (Laitin 2013: 44). Many of us constantly flit back-and-forth between argument and evidence in a determined quest to create persuasive explanations about the puzzles and problems we see both how and why immutable outcomes like democratization, violence, revolutions, elections, or urbanization happen in certain times and places. A very few of us articulate fully formed hypotheses from scratch before ever heading into the field for interviews, perusing historical archives, constructing an experimental design, or coding quantitative figures. Instead, we often make do with lesser hunches that consist of prior theoretical beliefs or explanatory conjectures that make some provisional claim about suspected causal relationships.[9]

---

[9] They sound, of course, far more sophisticated on our grant applications.

Furthermore, many of us modify these propositions midway through projects as insights strike organically. We may make untold methodological recalibrations, analytical revisions, and theoretical reinvention during process of research itself, and even add new cases, recapitulate existing variables, and reconceptualize causal mechanisms within the context of discovery. Such changes can occur suddenly as insights strike our mind when locked at an intellectual impasse, often when immersed in the field and poring over the same interviews, newspapers, ideas, and events. In this way, many of us treat concrete research progress not as a neat data set gleaned from laboratory procedures performed under controlled conditions, but rather as an oft-chaotic mass of information manifest in multiple forms—e.g., field notes and audio tapes, draft papers and rough chapters, e-mail summaries and questions, scribbles on xeroxed documents, and numbers in computational software. Such projects can be embarrassingly muddled until the very last "aha!" moment, when the structure of the explanatory argument suddenly crystallizes.

The close interplay of theory and data can mean that hypothesis generation and hypothesis testing blend together into a more iterated process of gradually distilling an explanatory proposition that can decisively account for the outcomes of interest. This resonates especially with those comparativists who take temporal factors into account, such as path dependency and sequentialism. We trace a path-dependent chain of causation by moving forward and backward across history, making sure each link connecting actors to events is sound; we may well conjecture that initial choices can deposit legacies that invariably shape long-term processes, but we have little idea about the scope, intensity, direction, sequence, and breadth of the argument unless we peek at history itself.

I call this interplay between theory and data "inductive iteration," but in any form, it manifests as a healthy mix of deduction and induction aiming to not simply confirm hypotheses but posit convincing and credible explanations. Figure 2 etches out a visual representation of this model of "inductive iteration" which can be harnessed with qualitative (especially case-oriented and historical) work, quantitative techniques, and game-theoretic modeling (Yom 2015a: 628).

The difference between (Hempelian) confirmation and (Lakatosian) explanation commands its own literature, with each playing a role in the broader philosophical enterprise of material inference (Miller 1987; Sklar 1999). However, here, I draw a critical distinction in what much area studies scholarship seeks. To be sure, many branches of empirical political science can adhere to the deductive template and its goal of confirmatory hypothesis-testing well, such as those that rely upon econometric modeling. Indeed, in some mainstream quantitative work, deviations from the deductive template are precisely how improprieties and abusive practices, such as data-mining and curve-fitting, begin. One reason is that the Rubin-Holland model of causality prevailing in much quantitative work requires counterfactual-conditional reasoning in which hypothetical models are tested against potential rather than observed outcomes.

Whether understood through qualitative methods or abstracted at a higher level of logic, however, many area specialists are concerned not only with hypothesis confirmation through the deductive template but also the construction of convincing explanations that account for interesting phenomena. For some, to confirm a causal
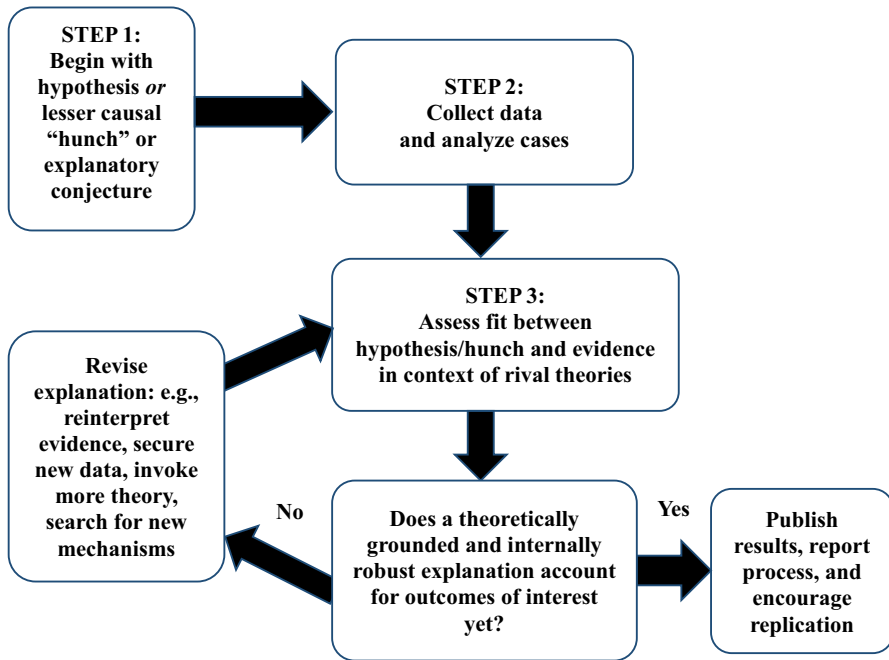
Fig. 2 Iterative research

hypothesis is to make the following statement: that given the evidence, believing that hypothesis to be true is more defensible than disbelief, though one must always be open to revision when new data are collected. For others, more totalizing and comprehensive explanations are required: persuasive explanations are those that identify all relevant conditions (including variables and mechanisms) that, in the cases under review, are collectively sufficient to bring about the outcome of interest.

For instance, we can hypothesize that a falling barometer marks the onset of an approaching cold front, but this is not a causal for why that cold front is approaching. The explanation would understanding many other moving parts, such as the movement of air masses, thermodynamic mechanisms, pressure differentials, and evaporative patterns. In comparative politics, to take Gregory Luebbert's work, a hypothesis claiming that alliances between urban socialists and the rural sector underlay the rise of Fascist governments would have been confirmed by the actual history of interwar Europe (1991). Yet, Luebbert's book is convincing only, because it engages in meticulous, process-oriented case studies to show in precise detail why certain constellations of urban parties, rural forces, and economic elites generated cascading patterns of mobilization, bargaining, and choices that ultimately resulted in diverging regime types across the European continent by the 1930s.

That inferential task of gradually building a theoretically guided explanations, often built upon the language of necessary and/or sufficient conditionality, drives a considerable degree of well-known area-specific projects, such as studies of ethnic conflict and peace in India (Varshney 2002), mass mobilization under communist

autocracies (Ekiert 1996), insurgency and regime change in South Africa and Central America (Wood 2000), civil society and political development in Italy (Putnam 1993), democratic consolidation in Southern Europe and Latin America (Linz and Stepan 1996), the legacies of colonialism in North Africa (Anderson 1986), and trajectories of industrialization in East Asia (Haggard 1990). The metric of success for these acclaimed studies was not the simplistic criterion of whether the authors confirmed their original hypothesis, but rather whether their overall theoretical arguments came across as credible—that is, able to account for every instantiation of the political outcome under question within all the cases, and thus persuasively answer the question of "why" in the context of all known rival explanations (Seager 1987).

4 Imagining Analytic Transparency: Testing Hypotheses

What do demands for greater analytic transparency mean in this context of inductive iteration and explanatory argumentation, which many area specialists carry out? For one, it connotes the belief that political science as a *science* is more comparable to, say, psychology or biology rather than anthropology and history, and that research we should conduct would follow the deductive template of hypothesis-testing. For another, applying statements calling for greater analytic transparency, such as those made by Gary King—"the only way to understand and evaluate an empirical analysis fully is to know the exact process by which the data were generated and the analysis produced (1995: 444)—to area-specific projects is far from simple.

Analytical transparency deals with the crux of the research process, the critical phase in which researchers realize whether they can churn out publishable conclusions. Yet, what does it fundamentally mean to divulge the exact process by which analyses are made and causal inferences rendered? It is here that the incongruity of comparing comparative politics research within geographic areas to the biological and physical sciences—a comparison that the language of DA–RT invites us to make—manifests. In other scientific arenas, analysis literally takes place within a controlled laboratory setting, or some confined space that contains all relevant data as well as the computational and mechanical equipment necessary to analyze it. It is difficult to bring one's work home in the way that most political scientists can, because it requires a very specific configuration of resources and materials that cannot be easily transported.

Take, for example, a recent and notable scientific finding in biology. In 2015, the University of Virginia team discovered a neurological link between the brain and immune system, which generations of medical textbooks had treated as unconnected entities (Louveau et al. 2015). The verification of meningeal lymphatic tissue carrying immune cells and small molecules between the brain and lymphatic nodes is startling given the assumption that all tissue structures in the body had been already mapped out. It also promises catalyze a radical rethinking of how neurological diseases originate and progress, such as Alzheimer's, autism, and multiple sclerosis. Now, the brain can be treated mechanistically, and its meningeal

lymphatic tissues treated as a causally relevant variable in neuroimmunology, with perhaps its malfunction even serving as a potential cause of major disease.

How did this significant result come about? When mounting the meninges of mice on a single slide by affixing the membrane to the skullcap, the medical team noticed vessel-like patterns in the distribution of immune cells. Excitedly hypothesizing that these could be lymphatic vessels, researchers decided to put their proposition under fire by testing for the presence of lymphatic vessels with new mice and human brain tissue. An exhaustive array of tests and formal protocols followed—dissections, incubations and buffering, antibody injections, staining, image analysis including statistical diagnostics, and both electron and multiphoton microscopies. Each of these hypothesis tests existed in the most literal sense; they were timestamped, logged, and recorded. More to the point, each step of this deductive process required a physical procedure involving resource-intensive inputs and externally perceivable results, with the object of study (brain tissue) analyzable only within the lab.

This example highlights some irreducible differences between many sciences and the study of politics. It is easy to show analytic transparency when hypothesis testing is spatially confined to a laboratory and each test consumes an actual moment in time. By contrast, it is extremely difficult to square analytic transparency with area specialization within comparative politics because of its iterative nature. First, in an ontological sense, many comparativists do not merely "test" hypotheses in piecemeal sense. Not only is area-specific research often observational in nature (though experimental work is rapidly growing), but tools of analysis such as process-tracing within case studies do not unfold as one-off explorations. In a typical scenario where, say, a Russian specialist wishes to show that economic liberalization in the 2000s came about due to the institutional legacies of elite alliances forged a decade earlier, a process-tracing narrative might begin by highlighting the organizational and political landscape of post-Soviet Russia, then trace the social and provincial origins of different elite configurations, present evidence of elite-based meetings, partnerships, and contracts, illustrate the causal mechanisms that translated these elite linkages into new policy preferences, and otherwise, present all contextual details that might help readers to evaluate the explanation in light of rival explanations, such as the counterclaim that economic liberalization came about due to external pressures instead.

If the Russian specialist were to justify this project as testing the hypothesis that post-Soviet elite alliances caused economic liberalization, the actual "test" comes in the constructed and prosaic landscape of the case study, as the reader absorbs each word. The test is figurative, and comes across as an idealized image of a tantalizing proposition clashing with fascinating evidence, then walking away the confirmed winner. In real life, the Russian specialist most certainly does *not* sit down with a mass of data and, in a single writing session, plot out each link of the process-tracing narrative blindly, and then only after the writing is done step back and assess whether the narrative matches the original hypothesis plotted months or years earlier. [Tasha Fairfield, however, has devised a fascinating way using Bayesian logic to formalize process-tracing by assigning matched probabilities to every causal step in the narrative (Fairfield 2015)]. The case study is written over

time, revised repeatedly; and every time the author sits down to extend or edit the document, and the hypothesis is being continually "tested" in his mind as he attempts to make sense of everything. Only in the end, when the case study is finished in whatever form—an article, a book chapter, a thesis— can the scholar present the actual case study to the outside work as something representing a "test," and even then many are more likely to see that case study as something else: the construction of a persuasive causal explanation.

Put another way, in the laboratory, the hypothesis test occurs in a physical setting can be meticulously logged, mapped, and recorded. In much comparative political scholarship, the hypothesis test exists in two forms: it occurs many times, constantly and repeatedly, in the mind of the scholar as he works through theory and evidence; and it occurs once in the mind of the reader, as she imagines the argument being subjected to the implications of the data as presented.

A related point is that iteration itself is messy. If area studies specialists must divulge every analytical step taken during the construction of an explanation—that is, every oscillation between argument and data, every relevant thought that played some role in the intellectual chain of analysis that concluded in the published explanation—the result might be a hideously unreadable log. For instance, Ph.D. candidates writing their dissertations within comparative politics seldom see their projects as an austere matter of hypothesis confirmation; many (if not most) projects are heavily revised, recalibrated, and even reinvented over the gritty years of graduate training. Yet, academic advisers prefer to not read doctoral theses that read like an orrery of the errors made before the writer stumbled upon a convincing answer. Typically, these stories are best told over drinks at APSA hotel bars, not at the dissertation defense.

To take another example, political scientists seldom shrug their shoulders and declare a grandiose project a failure when contradictory results emerge (such as when field-based interviews clash with our priors); nor do they treat less-than-optimal choices made at the beginning of research (inappropriate variables, misinformed indicators, misshapen concepts) as givens—they improve, revise, and change as they inductively learn more about. I will take my book, *From Resilience to Revolution: How Foreign Interventions Destabilize the Middle East*, as an example (Yom 2015b). *From Resilience to Revolution* locates itself within the comparative politics of the Middle East. It advances a sweeping theoretical proposition: when outside powers, like the US, intervened with foreign aid and military support to help dictators consolidate power over opposition early in their post-colonial tenures, they tended to sabotage the prospects for future stability by encouraging those autocrats to mobilize very narrow coalitions, become reliant upon repression, and generally shrug off rule-maximizing laws. Thus, levels of stability exhibited in post-colonial autocracies in the Middle East can be explained by how dollops of foreign aid (or lack thereof) affected the earliest policies of these governments.

This was no simplistic exercise in hypothesis confirmation. My initial belief was precisely the opposite—the more foreign aid, the more stable countries would become over time. Historical evidence from cases like Iran, however, forced me to rethink that founding conjecture. Elite interviewees in Jordan who linked US aid

programs with destructive effects decisively shattered my confidence in the proposition. One year into writing, I should have declared the hypothesis disconfirmed, but I did no such thing (not least because of my overseas commitments and grant obligations!). I continued traveling to other countries, such as Kuwait and Tunisia, back to Jordan, side-trips to Pakistan, Bahrain, Lebanon, the UAE, and Morocco, and over those years revitalized my argument with a fresh theoretical approach. I read new literature, including work in international hierarchy and cliency, and asked new questions with every encounter with new data in the form of interviews, archives, government documentation, and declassified materials. The thesis advanced in the book was the product of this meticulous and iterative process of both theory-building and theory-testing, of generating a persuasive causal explanation for trajectories of authoritarian durability in the Middle East that held up to rival arguments like those emphasizing oil or economics.

With due respect to analytic transparency, what part of the "exact process by which the data were generated and the analysis produced" should be explicated? The expected version traffics in vocabulary familiar to mainstream comparativists: that immersed in theories of international cliency, post-colonial state formation, and authoritarian durability, I proposed a daring new hypothesis in which great power assistance reduced the long-term stability of recipient dictatorships; and to prove it, I "went" into three cases (Iran, Jordan, Kuwait), observed that data corroborated the proposition, displayed that evidence through comparative detailed process-tracing, and thus could conclude the project as a successful exercise in the rigorous testing of hypotheses.

The second, and more truthful version, proceeds like this. I started out boldly but wrong; indeed, the initial hypothesis dissolved within the first 6 months of research. I considered new variables (leadership choices, external pressure, ethnic conflicts) that were never in the original research or grant proposals that help explain how foreign aid translated into domestic outcomes on the ground. I began asking different questions when interviewing elites, looking for different kinds of evidence in historical archives, and returned to read many theoretical scholarships that had never influenced the original research design. My case studies were not inscribed in a single and irrevocable time period: rather, I wrote, rethought, and rewrote again these analyses innumerable times in fits and spurts. Sometimes, these recalibrations reflected new insights and ideas that emerged in the process of writing itself. Other times, inspiration came from book editors, peer reviewers, and friends who read this work prior to publication. In short, I wrote this book in the same way that many of the great *doyen* of classic comparative analyses did: not in a matter-of-factly manner of dispassionately testing a hypothesis, but in an organic and open-ended way consuming several years of life in which both induction and deduction mixed together in an ongoing intellectual dialog driven forward by serendipity as much as science. [Gerardo Munck and Richard Snyder's interviews with the *doyen* of comparative politics say as much. In their volume, scholar after scholar (e.g., Adam Przeworski, David Collier, Barrington Moore, Jr.) recounts the real story behind their major *ouvres*, which differ greatly from published accounts of how their research projects proceeded (Munck and Snyder 2007)].

Analytic transparency means divulging the precise process by which conclusions are generated from data, but the first rather than second version of this research practice remains the preferred one for the academic market. Therein, lays the problem. Moving forward and heeding the call for analytic transparency, comparativists must either lie and make sure the real story never sees the light of day, or else be exposed as subjective, unscientific, backwards, or even anti-positivist for violating the deductive template. Resolving this dilemma should not even be a question; it is impossible to reconcile these two sides, the idealized version of hyperbolic laboratory science versus the gritty enterprise of daily research and writing. Yet so long as analytic transparency is cherished as a pillar of good research practice, scholars will inevitably bounce between these two poles.

Another detriment is that demands for full analytic transparency may ironically *deter* replication, the very benchmark that DA–RT principles are supposed to enact. If an author's scholarly outcome is to be replicated by third-party researchers to verify its robustness, third-party researchers may very well back away if they are fully told the mammoth quantity of resources, extreme amounts of time, litany of intellectual detours, and other intricate steps needed to generate the precise same result using the same data and cases.[10] In addition, it is unlikely they could reproduce the same exact procedures.

5 Conclusion

This essay has outlined several points of concern regarding the current climate of imposing greater transparency, in particular analytic transparency, within political science and especially area-specific venues of comparative politics. The origins of DA–RT may reflect other scientific disciplines, but its implementation within its Western core has proceeded rapidly in line with the biological and physical sciences. The impetus for greater transparency, however, rests upon an extremely restrictive template of purely deductive procedures that many political scientists cannot follow in the real world. Area-specific scholarship, with its requirements to gain deep knowledge and expertise over specific intersections of geography and time, is seldom elegant in practice. It involves a great bit of back-and-forth movement between theory and evidence within the context of intellectual discovery. Often times, there is no identifiable hypothesis test as can be discerned in the natural sciences within laboratory-based settings. For many scholars in the field, the goal is not so much to confirm a barren proposition as it is to construct a convincing and credible explanation that accounts for observed outcomes. In creating those explanations, comparativists make mistakes repeatedly but adapt to them. Arguments are changed, cases are reevaluated, and information dissected further. Missteps are as much a part of this process as are right guesses and accurate conjectures.

If area specialization within comparative politics maintains its current vibrant place, with a few constraints on the eclectic practice of its adherents, then it will

---

[10] I am indebted to Philippe Schmitter for reminding me of this point.

remain a central part of the subfield for generations to come. However, if its practice becomes subsumed under the all-encompassing mantra of transparency advocated by legitimate science, then its future looks gloomier. One of these two possibilities will become reality within the next decade.

References

Anderson, Lisa. 1986. *The state and social transformation in Tunisia and Libya, 1830–1980*. Princeton: Princeton University Press.

Bates, Roberts. 1997. Area studies and political science: rupture and possible synthesis. *Africa Today* 44: 123–132.

Clarke, Kevin, and David Primo. 2012. *A model discipline: Political science and the logic of representation*. New York: Oxford University Press.

Deer, Brian. 2011. How the case against the MMR vaccine was fixed. *BMJ* 342: c5347.

Edmunds, Scott. 2013. Peering into peer-review at Gigascience. *Gigascience* 2: 1.

Ekiert, Grzegorz. 1996. *The state against society: Political crises and their aftermath in east central Europe*. Princeton: Princeton University Press.

Fairfield, Tasha. 2015. Reflections on analytic transparency in process tracing research. *Newsletter of the American Political Science Association organized section for qualitative and multi-method research* 13: 47–51.

Fanelli, Daniele. 2012. Negative results are disappearing from most disciplines and countries. *Scientometrics* 90: 891–904.

Gerber, Alan, and Neil Mahotra. 2008. Do statistical reporting standards affect what is published? Publication bias in two leading political science journals. *Quarterly Journal of Political Science* 3: 313–326.

Haggard, Stephan. 1990. *Pathways from the periphery: The politics of growth in the newly industrializing countries*. Ithaca: Cornell University Press.

Hanson, Stephen. 2009. The contribution of area studies. In *Sage Handbook of Comparative Politics*, ed. Todd Landman, and Neil Robinson, 159–174. Thousand Oaks: Sage Publications.

Hochschild, Jennifer, David Lake, and Rodney Hero. 2015. Data access and research transparency initiative. PSNow Blog. http://www.politicalsciencenow.com/data-access-and-research-transparency-initiative-da-rt. Accessed 2 April 2016.

Ioannidis, John. 2005. Why most published research findings are false. *PLoS Medicine* 2: 696–701.

Ioannidis, John. 2008. Why most discovered true associations are inflated. *Epidemiology* 19: 640–648.

JAMA. 2016. Sharing clinical trial data: a proposal from the international committee of medical journal editors. *JAMA* 315: 467–468.

King, Gary. 1995. Replication, replication. *PS. Political Science and Politics* 28: 443–449.

LaCour, Michael, and Donald Green. 2014. When contact changes minds: an experiment on transmission of support for gay equality. *Science* 346: 1366–1369.

Laitin, David. 2013. Fisheries management. *Political Analysis* 21: 42–47.

Linz, Juan, and Alfred Stepan. 1996. *Problems of democratic transition and consolidation: Southern Europe, South America, and post-communist Europe*. Baltimore: The Johns Hopkins University Press.

Louveau, Antoine, et al. 2015. Structural and functional features of central nervous system lymphatic vessels. *Nature* 523: 337–341.

Luebbert, Gregory. 1991. *Liberalism, fascism, or social democracy: Social classes and the political origins of regimes in interwar Europe*, 1991. New York: Oxford University Press.

Masicampo, E.J., and D.R. Lalande. 2012. A peculiar prevalence of p-values just below.05. *Quarterly Journal of Experimental Psychology* 65: 2271–2279.

McDermott, Rose. 2010. Data collection and collaboration. *PS. Political Science and Politics* 43: 15–16.

McNutt, Marcia. 2014. Reproducibility. *Science* 343: 229.

Miller, Richard. 1987. *Fact and method: Explanation, confirmation, and reality in the natural and social sciences*. Princeton: Princeton University Press.

Moravcsik, Andrew. 2010. Active citation: a precondition for replicable qualitative research. *PS Political Science and Politics* 43: 29–35.

Munck, Gerardo, and Richard Snyder (eds.). 2007. *Passion, craft, and method in comparative politics*. Baltimore: The Johns Hopkins University Press.

Nature. 2013. Reducing our irreproducibility. *Nature* 496: 398.

Open Science Collaboration. 2015. Estimating the reproducibility of psychological science. *Science* 349: 943.

Prin, Florian, Thomas Schlange, and Khusru Asadullah. 2011. Believe it or not: How much can we rely on published data on potential drug targets? *National Reviews Drug Discovery* 10: 712.

Pusztai, Lajos, Christos Hatzis, and Fabrice Andre. 2013. Reproducibility of research and preclinical validation: problems and solutions. *Nature Reviews Clinical Oncology* 10: 720–724.

Putnam, Robert. 1993. *Making democracy work: Civic traditions in modern Italy*. Princeton: Princeton University Press.

Rosenthal, Robert. 1979. The file drawer problem and tolerance for null results. *Psychological Bulletin* 86: 638–641.

Seager, William. 1987. Credibility, confirmation, and explanation. *British Journal for the Philosophy of Science* 38: 301–317.

Sil, Rudra. 2009. Area studies, comparative politics, and the role of cross-regional small-n comparison. *Newsletter of the American Political Science Association organized section for qualitative and multi-method research* 7: 26–32.

Simmons, Joseph, Leif Nelson, and Uri Simonsohn. 2011. False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science* 22: 1359–1366.

Sklar, Lawrence. 1999. The content of science, the methodology of science, and hempel's models of explanation and confirmation. *Philosophical Studies* 94: 21–34.

Sterling, Theodore. 1959. Publication decisions and their possible effects on inferences drawn from tests of significance—or vice versa. *Journal of the American Statistical Association* 54: 30–34.

Szanton, David. 2004. The origin, nature, and challenges of area studies in the United States. In *The politics of knowledge: Area studies and the disciplines*, ed. David Szanton, 1–33. Berkeley: University of California Press.

Varshney, Ashutosh. 2002. *Ethnic conflict and civic life: Hindus and Muslims in India*. New Haven: Yale University Press.

Wainstock, Dan. 2014. Transparency. *Developmental Cell* 30: 489.

Waldner, David. 2007. Transforming inferences into explanations: Lessons from the study of mass extinctions. In *Theory and Evidence in Comparative Politics and International Relations*, eds. Richard Ned Lebow and Mark Irving Lichbach, 145-76. New York: Palgrave Macmillan.

Wood, Elisabeth Jean. 2000. *Forging democracy from below: Insurgent transitions in South Africa and El Salvador*. New York: Cambridge University Press.

Yom, Sean. 2015a. From methodology to practice: Inductive iteration in comparative research. *Comparative Political Studies* 48: 616–644.

Yom, Sean. 2015b. *From resilience to revolution: How foreign interventions destabilize the Middle East*. New York: Columbia University Press.

Sean Yom Sean Yom is Associate Professor of Political Science at Temple University in Philadelphia, USA. He specializes in the politics and development of post-colonial states in the Middle East. He is the author of From Resilience to Revolution: How Foreign Interventions Destabilize the Middle East (Columbia University Press, 2016) and various articles appearing in Comparative Political Studies, Studies in Comparative International Development, Journal of Democracy, and other journals.