

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI[®]

**THINKING *INSIDE* THE (BLACK) BOX: A
SCIENTIFIC REALIST-BASED APPROACH
TO CAUSAL INFERENCE IN POLITICAL
SCIENCE**

by

Daniel Meredith Schwartz

A thesis submitted in conformity with the requirements

for the degree of Doctor of Philosophy

Graduate Department of Political Science

University of Toronto

© Copyright by Daniel Meredith Schwartz 2005



Library and
Archives Canada

Bibliothèque et
Archives Canada

0-494-07823-5

Published Heritage
Branch

Direction du
Patrimoine de l'édition

395 Wellington Street
Ottawa ON K1A 0N4
Canada

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file *Votre référence*

ISBN:

Our file *Notre référence*

ISBN:

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.


Canada

To my family
For all their love and support

To know truly is to know by causes – Francis Bacon

Happy are those who know the causes of things – Virgil

The law of causality, I believe, like much that passes muster among philosophers, is a relic of a bygone era, surviving, like the monarchy, only because it is erroneously supposed to do no harm – Bertrand Russell

Title: Thinking *Inside* The (Black) Box: A Scientific Realist-Based Approach To Causal Inference In Political Science

Degree and Year: Doctor of Philosophy, 2005

Name: Daniel Meredith Schwartz

Department: Political Science

University: University of Toronto

Abstract

Like all social science disciplines, much of political science is devoted to making causal inference. But the quest to make causal inference in the discipline is undermined by dominant and long-standing philosophical traditions, which hold that causation is not a genuinely meaningful concept. Thus, in the tradition of *Humean empiricism*, causal relationships in the discipline are sometimes treated as unexplained correlations; and when mechanisms are offered to explain correlations, it is often done in the tradition of *instrumentalism empiricism*: mechanisms are relatively simplistic and treated as fictions that do not possess genuine explanatory power.

In short, political scientists often take a *black box* approach to causation, emphasizing empirical regularities over theoretical mechanisms. When political scientists do attempt to identify accurate and detailed theoretical mechanisms that reflect the complexity of real world social processes — when the black box is opened and explored — their work is often conducted without regard to correlational evidence. This too is a result of the

powerful philosophical traditions that favour regularities over mechanisms: correlational methods are often considered a distinct sphere of activity, separate from, and superior to, non-correlational methods that are geared primarily toward identifying social mechanisms.

This thesis attempts to develop a philosophical basis that complements, rather than clashes with, the desire to make causal inference. Building on *scientific realist* literature, I develop a unique approach to making causal inference: an approach applicable to political science and, more generally, the social sciences. The social scientific realist (SSR) approach that I develop holds that theoretical mechanisms are indispensable to establishing causation: thinking *inside* the black box is critical. But my approach also holds that theoretical mechanisms cannot be firmly established in isolation from correlational evidence: to think seriously *inside* the black box, it is necessary to embrace rather than eschew correlational methods.

Acknowledgements

This thesis would not have been possible without the dedication of my supervisor, Tad Homer-Dixon. Tad spent many hours reading through numerous versions of my thesis, and throughout the process offered me insightful and invaluable critique and support.

My other thesis committee members also played a crucial role in developing the ideas contained in this thesis. I benefited greatly from the guidance and support of Don Forbes and Jim Brown (Philosophy). Andrew Bennett (Georgetown) and David Welch also provided helpful and encouraging comments.

Finally, I would like to acknowledge the support of my entire family throughout the lengthy process of completing this thesis. In particular, I would like to thank my parents, Myrna and Irwin Schwartz, for never giving up hope and for their willingness to help whenever and however needed. I would like to also especially thank my wife, Celina Weigensberg, for her unwavering encouragement and advice. In addition to her loving support, Celina also helped me think through countless substantive issues relating to philosophical and methodological matters. Celina is the silent co-author of this thesis.

Table of Contents

Abstract...iv

Acknowledgements...vi

Abbreviations...xii

Preface...xiii

Chapter I: Introduction...1

i. Core Components...4

ii. Outline...13

iii. Caveats and Disclaimers...14

**Part I: The Black Box Approach to Causation and its Scientific Realist
Challengers...19**

Chapter II: Hume and the Black Box Approach to Causation...20

i. The Black Box Approach...21

ii. Hume's Philosophy of Causation...24

iii. Conclusions: Looking Ahead...32

Chapter III: The Causal Mechanism Approach...34

i. Causal Mechanism in the Natural Sciences...36

ii. Causal Mechanism in the Social Sciences...45

iii. Conclusions: Pragmatic Implications...51

Chapter IV: The Causal Power Approach...43

i. Causal Power in the Natural Sciences...44

ii. Causal Power in the Social Sciences...68

iii. Conclusions: Pragmatic Implications...72

Part II: The Social Scientific Realist Approach...74

Chapter V: Why Causal Mechanism Trumps Causal Power...75

i. Answering Causal Questions...76

ii. The Quest for Genuine Universality...85

iii. Conclusions: A Deeper Look at Mechanisms...87

Chapter VI: Physicalness and Agency...89

i. Defining Mechanism...94

ii. Physicalness, Process-Tracing, and Regularities...96

iii. Agency, Rational Choice Modeling, and Regularities...99

iv. Conclusions: Intuitiveness and Acceptance...112

Chapter VII: Intentionality...114

i. The Method of Interpretation...115

ii. Merging Interpretation and Causal Analysis...116

iii. Interpretation and Mechanisms in Political Science...127

iv. Conclusions: Cross-Fertilization and Testing Mechanisms...143

Chapter VIII: The Reality of Ideas...	147
i. Scientific Realism's Ontology and its Application to the Social Sciences...	149
ii. The External Reality of Ideas...	152
iii. Conclusions: Application to the Three Ontological Properties of Mechanisms...	159
Chapter IX: Mechanisms and the End of Explanation...	161
i. Salmon and the End of Explanation...	165
ii. Aronson and the End of Explanation...	169
iii. Conclusions: Social Mechanisms and the End of Explanation...	172
Chapter X: The Epistemological and Ontological Status of Mechanisms...	175
i. Four Positions on Causation...	176
ii. Causal Inference: The Difference that Scientific Realism Makes...	200
iii. Causation and Epistemic Thresholds...	214
iv. Conclusions: A Summary of The SSR Approach to Causation...	217
Part III: Causal Inference in Political Science...	219
Chapter XI: DSI and Causal Inference...	220
i. DSI and Humean Empiricism...	222
ii. The Importance of Theoretical Mechanisms in DSI...	226

iii. Theoretical Mechanisms as per Instrumentalism Empiricism or Scientific Realism?...234

iv. Conclusions: Tension in Political Science...238

Chapter XII: Causal Inference in Democratic Peace Research...240

i. The Democratic Peace Debate...241

ii. The Importance of Theoretical Mechanisms in Democratic Peace Research...246

iii. Scientific Realism or Instrumentalism Empiricism in the Democratic Peace Literature?...258

iv. Conclusions: Methodological Prescriptions and Actual Research...265

Chapter XIII: The SSR Approach to Causal Inference as a Basis for Progress in Political Science...267

i. Training Practitioners...268

ii. Epistemic Unity...273

iii. Policy Prescriptions...293

iv. Conclusions: The Big Payoff...294

Chapter XIV: Conclusions...296

i. Summary...296

ii. Value Added to Existing Scientific Realist Approaches in the Social Sciences...301

iii. Potential Objections...305

iv. Broad Implications...317

v. Concluding Remarks...320

References...321

Figure 1...354

Table 1...355

Table 2...356

Table 3...357

Abbreviations

AfC argument from coincidence

DSI *Designing Social Inquiry*, by Gary King, Robert Keohane, and Sidney Verba

HE Humean empiricism or Humean empiricist

HR Humean realism or Humean realist

IE instrumentalism empiricism or instrumentalist empiricist

SR scientific realism or scientific realist

SSR social scientific realism or social scientific realist

Preface

The impetus for writing this thesis was a personal academic-related experience in 2001-2002 at the University of Toronto. At the time, I was peripherally involved in a large research project being conducted by the *Toronto Group* – led by principal investigator, Tad Homer-Dixon – on the causal relationship between environmental scarcity and violent conflict. The Toronto Group’s environmental scarcity research had recently come under fire from several critics on both theoretical and methodological grounds, most notably in a *Journal of Peace Research* article by Nils Petter Gleditsch (2001). Homer-Dixon requested that I consider the methodological criticisms voiced by Gleditsch and help respond wherever appropriate.

Gleditsch’s criticisms centered on the nature and value of case-study methodologies – process-tracing, in particular – employed by the Toronto Group and most other environment-conflict researchers. Gleditsch dismissed several key findings of the Toronto Group, contending that process-tracing generates “unsystematic” evidence and violate rules of quasi-experimental, correlational-based inference.

In researching and analyzing Gleditsch’s critique, I came to realize that his methodological approach typified conventional mainstream political science whereby correlations are the cornerstone of causal inference. But I found this correlational-based approach to causation puzzling and paradoxical: Hume taught us that correlations cannot underpin a *genuine* belief in causation; and yet mainstream political scientists such as Gleditsch were continuing to cling to their belief in causation, while simultaneously

insisting on the primacy of correlational-based evidence for assessing causal relationships.

This paradox of causal reasoning is evident in numerous methodological tracts in the discipline of political science (e.g., see Przeworski and Teune 1970), but perhaps never so prominently and so importantly as in King, Keohane, and Verba's (1994) *Designing Social Inquiry* (DSI). DSI has become nothing short of a methodological manifesto for the discipline of political science, despite the fact that it embodies the paradox of causal reasoning that I identify. DSI steadfastly and categorically refuses to countenance "noncausal explanations," and yet its formal definition of causation is built entirely on correlational principles.

In responding to Gleditsch's critique of the Toronto Group's work (Schwartz et al. 2001), I drew on arguments from a loose-knit group of social scientists – self-described as *scientific realists* – who argue that, rather than abandon the quest for causation, social scientists should supplement quantitative-based correlational evidence with qualitative-based evidence of causal mechanisms. In short, following these scientific realists (in particular, Alexander George and Andrew Bennett), I argued that causation amounts to regularities supplemented by mechanisms.

This thesis represents a refinement and elaboration of the ideas first presented in my response to Gleditsch's methodological critique. Although many of the ideas in Schwartz et al. (2001) have undergone significant transformation in this thesis, I remain resolute in

my underlying message: resolving the aforementioned paradox of causal reasoning requires a philosophical and methodological toolkit that makes causal analysis something more than mere correlational analysis.

Chapter I: Introduction

Like all social science disciplines, much of political science is devoted to making causal inference. But approaches to making causal inference in the discipline are heavily influenced by powerful and long-standing philosophical traditions, which hold that causation is not a genuinely meaningful term. Following in the tradition of *Humean empiricism* (HE)¹, causation is sometimes treated as nothing more than an unexplained empirical regularity, or correlation.² When theoretical mechanisms³ are advocated or used to explain correlations, it is often done in the tradition of *instrumentalism empiricism* (IE)⁴: mechanisms are relatively simplistic and treated as fictions that do not possess genuine explanatory power. Both HE and IE favour a *black box* approach to causation, because both emphasize empirical regularities to the exclusion (HE), or near-exclusion (IE), of theoretical mechanisms.

The influence of HE and IE has also contributed to deep methodological divisions in the discipline of political science, further impeding efforts at making causal inference.

Researchers who employ *process-tracing*, *interpretation*, and *rational choice modeling* – methods that can help identify detailed and accurate theoretical mechanisms that reflect

¹ HE will be used to refer to “Humean empiricism” or “Humean empiricist,” depending on the grammatical context.

² In Chapter II we will see that unexplained counterfactuals are also relevant.

³ Theoretical mechanisms generally refer to unobservable objects, events, and processes.

⁴ IE will be used to refer to “instrumentalism empiricism” or “instrumentalist empiricist,” depending on the grammatical context.

real-world processes – often work in isolation from researchers who employ correlational methods. This is, in part, attributable to the lasting influence of HE and IE in the discipline, both of which consider correlational research a distinct and superior sphere of research activity.⁵

In short, the desire to make causal inference in political science is not complemented by a philosophical basis on which to do so. The discipline is thus beset by philosophical inconsistency. In the tradition of *scientific realism* (SR)⁶, practitioners are taught that causation is a genuinely meaningful term; but they are simultaneously schooled in methods and techniques influenced by HE and IE. These philosophical inconsistencies are evident in the discipline’s methodological prescriptions, which champion an incoherent mix of HE, IE, and SR. For instance, the inconsistencies are glaringly evident in the discipline’s leading methodological text: *Designing Social Inquiry* (DSI), by Gary King, Robert Keohane, and Sydney Verba. DSI rejects “noncausal explanations” and yet simultaneously and paradoxically embraces a Humean regularity-based definition of causation. To add to this philosophical disarray, DSI also advocates certain pragmatic techniques for increasing the N of small-N study that implicitly champion elements of

⁵ According to George and Bennett (2005, 10), “...[O]nly about two-thirds of the thirty top-ranked graduate programs in political science offer a dedicated graduate course in qualitative or case study methods, and only two of these departments require such a course. In contrast, all of the top thirty departments offer courses in statistics, and almost all of these departments require some training in statistics, often several courses.”

⁶ SR will be used to refer to “scientific realism” or “scientific realist,” depending on the grammatical context.

both IE and SR.⁷ These inconsistencies in the discipline's methodological prescriptions, in turn, manifest themselves in political science research. Although leading research in the discipline largely eschews HE, a philosophically inconsistent mix of IE and SR often underlie its methodological approaches to causal inference.

This thesis attempts to develop a unique *social scientific realist* (SSR)⁸ approach to causal inference that could help resolve these inconsistencies. The SSR approach entails thinking more seriously *inside* the black box. By focusing on the identification and testing of theoretically accurate and detailed mechanisms, the SSR approach might get us beyond HE and IE, which pay only lip service to causation. Moreover, this approach promises to help advance the discipline of political science by bridging methodological divisions between and amongst the discipline's quantitative and qualitative researchers.

⁷ This is not suggest, even for a moment, that DSI is not without virtue. It is widely agreed that its treatment of rules pertaining to statistical inference is exceptionally lucid. In addition, as I will argue in Chapter XI, some of its pragmatic advice on how to increase the N of a small-N study can be extremely useful. I also do not want to suggest that DSI are unaware of issues pertaining to causal mechanisms; only that their treatment of causal mechanisms is, in my view, misguided. It is worth noting on this front that in later publications, Robert Keohane seems to have a more favourable outlook on the importance of causal mechanisms. See, in particular, Keohane (1999). Although critiques of DSI already exist – for particularly powerful critiques, see Brady and Collier (2004) and George and Bennett (2005) – I believe that my particular critique is unique.

⁸ SSR will be used to refer to either “social scientific realism” or “social scientific realist,” depending on the grammatical context.

The SSR approach to mending methodological divisions in the discipline stands in stark contrast to that approach advocated by DSI. DSI attempts to turn qualitative research into quantitative research by applying the same logic of inference to the former as applies to the latter (George and Bennett 2005, 11). The SSR approach, by contrast, contends that qualitative research *as is* has inherent value: the goal of the SSR approach is not to turn ideographic researchers into statisticians, but rather to encourage integration of ideographic and statistical research.⁹ In addition to helping mend methodological cleavages, I will argue that the SSR approach might help to produce better-trained practitioners in the discipline and generate more precise information for policymakers.

i. Core Components

This thesis consists of two core components. In the first, I develop the SSR approach to causal inference. In the second, I examine the “state of the art” of causal inference in political science and suggest how the SSR approach can advance the discipline.

The SSR Approach to Causal Inference

The SSR approach to causation focuses on the development and testing of theoretical mechanisms; it contrasts with social scientific approaches to causal inference based on HE and IE.¹⁰ Causal inference requires that correlations are explained by mechanisms (in

⁹ This leaves aside for the moment the issue of rational choice modeling, with which DSI does not deal.

¹⁰ In Chapter X we will see that the SSR approach also contrasts with Humean realism (HR).

contrast to HE) and that the unobservable entities and processes that constitute mechanisms are *literally true* and not mere *useful fictions* (in contrast to IE).

The SSR approach holds that mechanisms of social causation consist of three distinct ontological properties: physicalness, agency, and intentionality. Each of these three distinct ontological properties can be identified by three corresponding social scientific methods: namely, process-tracing, rational choice modeling, and interpretation. Each of these methods, in turn, can generate findings that explain regularities in a unique fashion. Process-tracing generates findings that explain regularities as a series of events or phenomena linked together in time and space. Rational choice modeling generates findings that explain regularities as the aggregation or strategic interaction of individual preferences. Interpretation generates findings that identify the intersubjective meanings that are constitutive of regularities.

Although the SSR approach stresses the importance of mechanisms, it does not do so to the exclusion of correlational evidence: in fact, mechanisms and regularities are equally important to the SSR approach. Process-tracing, rational choice modeling, and interpretation, can identify accurate and detailed theoretical mechanisms that reflect the complexity of real world phenomena; but these methods provide only partial tests of the mechanisms that they identify. To make causal inference, the SSR approach requires that the mechanisms are tested further in order to demonstrate that they operate as theorized.

In addition, because “what does not exist cannot cause” (Miller 1987, 382)¹¹, the SSR approach requires that the unobservable entities and process that constitute mechanisms are tested further in order to demonstrate that they are real, or literally true.

The *argument from coincidence* (AfC), a core feature of SR, provides the logic necessary to test in a more robust manner the mechanisms identified by process-tracing, rational choice modeling, and interpretation. By the logic of the AfC, observable implications are drawn from a theoretical mechanism, ideally pointing to a wide range of empirical domains. The mechanism is then tested against this wide range of empirical phenomena. Generally speaking, the more empirical phenomena that a mechanism can predict (future) and retrodict (past), the more likely it is that the mechanism is operating as theorized and is literally true.

The SSR approach to causal inference contrasts sharply with an approach based on HE. Whereas SR holds that mechanisms are relevant to causal explanations, and the unobservable entities and processes that comprise mechanisms have a truth value, HE holds neither belief. Causation, for HE, to the extent that the term has any meaning, consists solely of unexplained correlations. HE holds that the best we can do to move from correlation to causation is to focus our research efforts on the independent variable (IV) and dependent variable (DV) in question. Generally, this takes the form of adding

¹¹ As will be seen in my discussion of SR in Chapter VIII, it is actually more accurate to say, “what does not have a truth value, cannot cause” rather than to say, as Miller does, “what does not exist, cannot cause.”

observations to an existing correlation and controlling for as many potentially confounding factors as possible. Because there is no theoretical mechanism available to explain the correlation, however, even the most robust correlations cannot be considered causal, regardless of how many potentially confounding factors are controlled: hence the well-worn phrase, “correlation does not entail causation.”¹²

The difference between the SSR approach and an HE-based approach has important methodological implications for making causal inference. In contrast to HE, causal research with the SSR approach is driven by theoretical mechanisms. Causal inference is made by deriving a range of observable implications from a theoretical mechanism, and then testing these observable implications – using an array of IVs and DVs from different empirical domains – against the empirical record. Thus, unlike with HE, researchers focus on *multiple* correlations (as well as non-correlational empirical evidence).

The SSR approach to causal inference also contrasts with an approach based on IE, although the contrast is much more nuanced. Like SR, IE draws on theoretical mechanisms to predict and retrodict regularities in a wide range of empirical domains. Theoretical mechanisms, according to IE, are relevant to scientific inquiry in the sense that they help predict and retrodict. Unlike SR, however, IE holds that the unobservable

¹² Philosophers of social science recognize that correlations are always subject to spuriousness because one can never be certain that all confounding factors have been controlled. However, it is equally true that spuriousness becomes progressively less of a risk as more potentially confounding factors are controlled.

entities and mechanisms that constitute mechanisms are merely useful fictions. IE rejects causation because theoretical mechanisms are neither literally true nor false and therefore have no genuine explanatory power (because, again, what does not exist cannot cause).

IE champions parsimony and generalizability: the simpler the theoretical mechanism, the more empirical domains to which it is portable, and hence the more empirical phenomena it can predict and retrodict. Like IE, the SSR approach holds that portability is a virtue. But in contrast to IE, the SSR approach is willing to sacrifice parsimony and generalizability on the altar of accuracy and reality: if they are to be more than useful fictions and have genuine explanatory power, theoretical mechanisms must reflect the complexity of the real world.

The difference between the SSR approach and an IE-based approach has important methodological implications in relation to the value of process-tracing and interpretation, as well to the manner by which rational choice models are constructed. For the SSR approach, process-tracing and interpretation are invaluable tools because they identify theoretical mechanisms that reflect the complexity of the real world. IE downplays the utility of process-tracing and interpretation; in fact, the complexity of the mechanisms derived from these case-study methods can be harmful to its twin goals of parsimony and generalizability.

With regard to rational choice modeling, the difference between the SSR approach and an IE-based approach has implications for how researchers treat core assumptions related to

rationality, self-interest, and methodological individualism. IE assumes that rationality is constant, whereas the SSR approach contextualizes rationality by limiting it to certain times and places. Similarly, IE assumes that rational actors are almost always self-interested, whereas the SSR approach recognizes that “other regarding” factors sometimes motivate rational choices. Finally, IE assumes that rational choices are not influenced by factors outside the individual level of analysis, whereas the SSR approach allows that “rules of the game” can be introduced exogenously.¹³ In short, whereas IE favours parsimony in rational choice modeling in order to maximize generalizability, the SSR approach favours accuracy and detail, recognizing that this might curtail generalizability.

Causal Inference in the Discipline of Political Science

The second core component of this thesis examines the “state of the art” of causal inference in political science, and suggests how the SSR approach could help move the discipline forward. To assess the discipline’s methodological prescriptions vis-à-vis causal inference, I examine *Designing Social Inquiry* (DSI) by Gary King, Robert Keohane, and Sydney Verba – widely considered the leading methodological text in the discipline. To assess how causal inference is made in political science research, I examine research on the *democratic peace hypothesis* – widely considered to be one of the discipline’s most mature research areas.

¹³ The distinction between “thick” and “thin” conceptions of rationality is sometimes said to be the basis for differing approaches to these three assumptions.

SRs typically assert that social scientists (including, political scientists) take a HE-based approach to causal inference. My findings only marginally support this assertion. With regard to methodological prescriptions, I find that political scientists are taught an inconsistent mix of HE, IE, and SR. DSI's formal definition of causation is explicitly based in HE. But it is ambiguous as to whether IE or SR underpins its set of methodological techniques for increasing the N of a small-N study. And DSI's explicit argument that causation is a genuinely meaningful term is consistent with SR. The typical SR assertion, that social scientists base causal inference solely on HE, is even further off the mark with respect to political science research. I find that democratic peace researchers implicitly base their efforts at making causal inference in both IE and SR, usually an uneasy and inconsistent blend of the two.

By adopting the SSR approach to causal inference, political science can evidently overcome the philosophical inconsistencies that characterize the discipline's methodological prescriptions and its actual research. Philosophical diversity can be virtuous, but not when expressed within single methodological texts (as is the case with DSI) or individual studies (as is sometimes the case with research in the democratic peace). Moreover, political science practitioners can overcome the philosophical inconsistency between their desire to make causal inference and their adoption of philosophical doctrines that eschew causation. Unlike typical SRs in the social sciences, I do not contend that political science requires a radical rethinking in order to adopt SR as its philosophical foundation. Political science is not simply characterized by HE, but rather, an incoherent mix of HE, IE, and SR. Adopting SR as a philosophical foundation

for the discipline may not be as onerous as some SRs in the social sciences seem to imply.

The SSR approach also promises to advance the discipline on at least three additional fronts. First, the SSR approach to causation helps “make sense” of DSI. DSI offers sound research design advice, but the clarity of its advice is obscured by the haziness of its philosophical framework. Once DSI’s philosophical inconsistencies are resolved in favour of SR, researchers can better understand and even extend some of its advice. Thus, the SSR approach can yield better-trained practitioners.

Second, the SSR approach suggests inherent compatibility between statistical analysis (on the one hand) and process-tracing, rational choice modeling, and interpretation (on the other hand). These latter three methods can identify accurate and detailed theoretical mechanisms that reflect real-world processes; and statistical analysis can test the observable implications derived from these methods. This compatibility can help bridge a nomothetic-ideographic gap in the discipline between those political scientists who advocate and practice statistical analysis, and those who advocate and practice either process-tracing or interpretation. In addition, this compatibility can help bridge a nomothetic-nomothetic gap in the discipline between those political scientists who advocate and practice statistical analysis and those who advocate and practice rational choice modeling.

The SSR approach can also help mend an additional nomothetic-ideographic gap in the discipline, between those political scientists who advocate and practice either process-tracing or interpretation and those political scientists who advocate and practice rational choice modeling. As noted, the SSR approach to rational choice treats rationality, self-interest, and methodological individualism, as context-dependent assumptions. Process-tracing and interpretation can provide this context: both methods can help determine the circumstances under which actors behave rationally, when actors are likely to make decisions based on narrow self-interest and when they are more likely to be motivated by “other regarding” factors, and when “rules of the game” might be imposed exogenously.

Finally, the SSR approach can help close an ideographic-ideographic gap between those political scientists who advocate and practice process-tracing and those who advocate and practice interpretation. Collaboration between these methods can identify more accurate and detailed theoretical mechanisms than would otherwise be the case. In brief, the SSR approach can advance political science by bringing a measure of epistemic unity to a discipline deeply divided along several methodological lines.

The third and final way that the SSR approach can help advance the discipline is by providing policymakers better advice. Many political scientists recognize that the goal of making causal inference is not an end in itself, but rather a means to equip policymakers with a better understanding of how political and social worlds function. Because the SSR approach emphasizes the development of accurate and detailed theoretical mechanisms

that reflect genuine political and social processes, it promises to give policymakers more precise information on which to act.

ii. Outline

Part I of this thesis (Chapters II through IV) outlines the context for the emergence of SR in the social sciences. Chapter II describes the black box approach to causation, and the related influence of David Hume's philosophy of causation. Chapters III and IV describe two SR-based approaches to causation that are in direct response to the black box approach described in Chapter II: the causal mechanism approach and the causal power approach.

Part II of this thesis (Chapters V through X) develops the SSR approach to causation. I begin in Chapter V by arguing that a SR approach to causation should be based on the concept of causal mechanism rather than the concept of causal power, because the latter is plagued with serious philosophical flaws. In Chapters VI and VII, I define "mechanism," and elaborate the three ontological properties of social mechanisms: physicalness, agency, and intentionality. I provide illustrations from social scientific research that demonstrate how each of these three ontological properties can be identified by three corresponding social scientific methods (namely, process-tracing, rational choice modeling, and interpretation) and how each of these methods can generate findings that explain regularities in a unique fashion. Chapter VII includes an extensive defense for the compatibility of interpretation and causal analysis.

Chapter VIII moves the analysis from the epistemological to the ontological. I contend that *realism* in the social sciences is possible, despite the fact that the social world (and, consequently social mechanisms) is comprised primarily of mind-dependent ideas.

Chapter IX makes a slight detour in the progression of my argument in order to examine another aspect of mechanisms. Contra some SRs, I argue that mechanisms are always in want of deeper explanation. The full implication of this argument is drawn out in Chapter X, where I argue that judgments about causation ultimately depend on epistemic communities. The main purpose of Chapter X, however, is to refine the debate over causation. I distinguish four philosophical doctrines – HE, IE, HR, and SR – based on differing approaches to the epistemological and ontological status of theoretical mechanisms, and describe the final but crucial features of the SSR approach.

Part III of this thesis (Chapters XI through XIII) examines the “state of the art” of causal inference in the discipline of political science. In Chapters XI and XII, I examine, respectively, DSI and the democratic peace debate. In Chapter XIII, I demonstrate how the SSR approach to causation can help advance the discipline. Chapter XIV concludes the thesis: I summarize my core arguments, describe how the SSR approach rectifies shortcomings in existing SR approaches to causation, anticipate potential objections to my arguments, and consider some broad implications of my analysis.

iii. Caveats and Disclaimers

The topic of causation is vast, variegated, and complex. The issues confronted in this thesis deal with a narrow slice of the literature on causation, and still the issues are

hugely involved and represent a formidable challenge. As Van Fraassen (1980, 1) points out, the debate between empiricists and realists over causation goes back at least to Aristotelian times.¹⁴ And the debate has surfaced in numerous forms since this time, such as in the dialogue about causal analysis in history between R.G. Collingwood, Carl Hempel, and William Dray (e.g., see Dray 1989), or the dialectic over causation in economics between formalists, positivists, and institutionalists (e.g., see Wilber and Harrison 1978). I will be gratified if I am able to make even a small contribution to the narrow slice of causation that I deal with in this thesis.

This thesis assumes that many social scientists want to make causal inference, and that causation without generalization is not possible: all other things being equal, and within certain limits of random error, relationships between variables should be constant across time and space. This is not meant to discount the possibility that causal relationships can be (and probably very often are) interactive and nonlinear, although complexity theorists might contend that the approach developed in this thesis unrealistically presupposes an ability on the part of social scientists to decompose causal relationships into component parts and control confounding influences.¹⁵ Further, my thesis assumes that prediction based on causal analysis is possible, although some complexity theorists might suggest

¹⁴ The debate between realists and empiricists in these early times was likely different in many respects from the debate that ensues in contemporary times.

¹⁵ Some philosophers of social science (e.g., Fay 1994 [1983], Louch 1966, and Strawson 1985) contend that generalizations are not possible because of “free will” and human reflexivity.

that interactivity and nonlinearities make prediction virtually unobtainable (Homer-Dixon 2000).¹⁶

In some sense my arguments are based on *naturalism* – the notion that if social science differs from natural science, it does so only in degree, not in kind (e.g., see Nagel 1961, 450-85). My focus on causation and generalization is in accordance with naturalism. However, the SSR approach recognizes that agency and intentionality makes the social sciences unique. I assume that human action fosters a measure of complexity in causal structures. Whereas some SRs in the natural sciences may correctly believe that parsimonious theories can represent real processes, the SSR approach assumes that simplistic and elegant theories are ill suited to represent reality in the social world.

This thesis utilizes research and examples from philosophy, the humanities, the natural sciences, and the social sciences. Some may find this diversity refreshing, others distracting. I can only say that literature was not chosen for its diversity per se, but rather because it helps to make my arguments. I try to focus wherever possible on political

¹⁶ George and Bennett (2005, 7, 9, 22, 231-62) contend that case study methods are particularly well suited to dealing with complexity, including “assessing complex causal relations such as path dependence, tipping points, multiple interactions effects, selection effects, disproportionate feedback loops, equifinality...and multifinality...” In fact, the authors argue that case study methods can be even better suited to this task than statistical methods, although their enthusiasm for case studies in this regard is carefully tempered (e.g., see p.116). For more on the limits of statistical methods vis-à-vis complexity, see Ragin (1987, 19-68).

science, and the social sciences more generally. But the nature of the topic requires material that takes us out of the realm of political science, traditionally understood.¹⁷

My arguments draw heavily on the philosophy of SR, which most often deals in the realm of the natural sciences. The SSR approach, however, is not applicable outside the social sciences. Although I assume that both natural and social scientists seek to make causal inference, and I draw on research and examples from the natural sciences, the SSR approach deals with characteristics unique to social and political worlds. Moreover, my approach to SR may break with traditional SR approaches. My argument for the tradeoff between detail and generalizability in theoretical mechanisms may be an example of such a break.

I use the terms, “regularity” and “correlation” synonymously throughout this thesis. More generally, I assume that HE does *not* require *invariant* relationships. Although this may be true of some HEs, most will allow for statistical or probabilistic relationships, where the probability of the effect given the cause is less than one, or $P(E/C) < 1$.

This thesis does not deal with the issue of *singular causation*, or events that occur only once, such as the mass extinction of dinosaurs 65 million years ago. But there may be reason to believe that singular causation can be treated in much the same manner as non-singular causation. Singular causal events require both regularities and mechanisms. We

¹⁷ It is not uncommon for philosophy of social science texts to stray far from the social sciences (e.g., see McKim and Tuner 1997, and Miller 1987).

infer that a meteorite could have extinguished the dinosaur species because we know from experience (this is the regularity part) that when large objects crash into other objects with great velocity and force, destruction ensues. Moreover, we have an understanding from our knowledge of biology, chemistry, and physics (this is the mechanism part) that once the meteorite did crash into Earth, certain physiological events would have unfolded that created, over time, an unlivable environment for the dinosaur species.

Finally, following SR, the SSR approach that I develop in this paper contends that we can make strong claims about unobservable entities and processes based on the logic of the AfC. In this sense, the SSR approach is anti-Humean and, in fact, claims to “defeat” Hume’s scepticism about unobservables (although it says little about Hume’s idealist scepticism about the observable world). But there is an important sense in which the SSR approach fails to defeat Hume’s analysis of causation. Hume refused to make truth claims because of his extremely high epistemological standards: Hume’s ontological commitments were consistent with his epistemological commitments. To the extent that the SSR approach defeats Hume, it does so only by adopting a lower epistemological standard than Hume would have allowed.

With these caveats and disclaimers in mind, I now turn to Part I of this thesis, which describes the black box approach to causation and its SR challengers.

Part I: The Black Box Approach to Causation and its Scientific Realist Challengers

SRs claim that the mainstream social scientific approach to causation – affiliated to statistical analysis, Mill’s methods, and Hempel’s models of explanation – champions the following formula: *causation = correlation*. Following Hume, this approach denies that causation is a genuinely meaningful term: the unobservable entities and processes that *explain* correlations are “unknowable,” and hence correlations are the single-best indicator of causation. The SR approach, conversely, holds that causation is a genuinely meaningful term because theoretical explanations are “knowable,” and can be invoked to explain correlations. The SR approach to causation champions the following formula: *causation = regularity + explanation*.

In the following three chapters, I outline these competing approaches to causation. Chapter II discusses the black box approach to causation and its affiliated methodologies. I contend that the SR understanding of the black box approach is misguided, and I hint at how the SSR approach will rectify this misunderstanding. Chapters III and IV discuss, respectively, two distinct (but often conflated) SR-based approaches to causation that have emerged in response to the black box approach: the *causal mechanism* approach and the *causal power* approach.

Chapter II: Hume and the Black Box Approach to Causation

One can only make sense of the causal mechanism and causal power approaches to causation, described respectively in Chapters III and IV, as a response to a Humean-based black box approach to causation. According to SRs, this Humean-based black box approach dominates mainstream social (and natural) science. In Part III of this thesis, I will argue that, if political science is typical of other social science disciplines, Humean causation is not quite as dominant as SRs claim. Moreover, the SR understanding of mainstream methodologies is sometimes misguided. Nevertheless, there is an important element of truth in the charge made by SRs, and even the SSR approach to causation that I develop in this thesis cannot be understood apart from the philosophical legacy of David Hume.

The first section of this chapter briefly describes what is meant by the black box approach to causation. The second section sketches Hume's philosophy of causation that underpins the black box approach. The conclusions point ahead to Parts II and III, where I refine what is meant by a black box approach to causation and argue that mainstream methodologies are not necessarily synonymous with such an approach.

i. The Black Box Approach

According to SRs, the black box approach to making causal inference in the social (and natural) sciences is founded on at least two related tenets.¹⁸ The first tenet holds that observable regularities, or correlations, are the sole indicator of a causal relationship. The second tenet holds that unobservable entities and processes have no legitimate role to play in explaining observable regularities. In short, proponents of the black box approach to causation are said to champion the following formula: *Causation = (observable) regularity*.¹⁹ This approach denies that causation is a genuinely meaningful term, because causation is reduced to nothing more than an unexplained regularity, or correlation (e.g., see Dessler 1991, Little 1991, Salmon 1989).²⁰

¹⁸ SRs vary on the extent to which they accuse mainstream political science of HE. George and Bennett (2005), for instance, do not necessarily suggest that only HE characterizes mainstream political science, although they do suggest that prominent methodological texts such as DSI seriously underemphasize the importance of mechanisms.

¹⁹ I say “observable” regularity, because as we shall see in Chapter X, HR holds a belief in unobservable regularities.

²⁰ Generally, SRs agree that theoretical explanations are not *entirely* eschewed by the black box approach to causation. Save for some data-mining techniques – whereby correlations between phenomena are discovered by systematically combing through masses of data in search of relationships – causal analysis generally begins by linking a phenomenon to be explained to an antecedent phenomenon by way of an explanation. For instance, an investigator looking for causes of famine might consider the relationship between agricultural production and famine because of the logical connection between the two: low agricultural production can engender high food prices, which in turn, can engender famine. Beyond initiating the investigation of a possible

SRs typically agree that mainstream methodologies, such as statistical analysis, Mill's methods, and Hempel's models of explanation, adhere to the two aforementioned tenets: causal inference with these methods is made with unexplained correlations.²¹ Statistical methods are evidently founded on the principle of correlation, and nothing in these methods technically requires a researcher to consider the reasons behind such correlations. This is true of nearly all statistical methods, from simple bivariate analysis to complex multivariate techniques. Mahoney (2001, 575) notes that: "Mainstream sociologists often infer causation through the use of quantitative techniques that depend, in one way or another, on the existence of bivariate correlations. Even sophisticated multivariate statistical methods that allow for the parceling of variables and the estimation of average net causal effects ultimately rely on such correlations."

Mill's *method of agreement* argues that if two or more instances of a phenomenon under investigation have only one of several antecedent phenomena in common, then the antecedent phenomenon in which all the instances agree is causally related to the phenomenon to be explained. Thus, the basis for making causal inference with Mill's

correlation, however, SRs claim that explanatory processes in mainstream causal analysis have little role to play (e.g., see Dessler 1989, and Little 1991).

²¹ Technically, some of these methods – such as Mill's *Method of Difference* – are based on counterfactual evidence. However, this distinction is generally glossed over in the SR literature because, like regularities, counterfactuals are left unexplained. For an in-depth discussion of counterfactual analysis in political science, see Tetlock and Belkin (1996).

method of agreement is an invariant relationship between two phenomena: or, in short, a regularity. Mill's method of agreement therefore adheres to the first aforementioned tenet of the black box approach. In addition, there is no question in the method of agreement as to *why* an invariant relationship might exist: therefore this method also adheres to the second tenet.

Or consider Hempel's *inductive statistical (I-S) model of explanation*: a particular phenomenon to be explained (O_i) is expected with "high logical, or inductive, probability," in light of particular events (F_i), and certain statistical laws whereby the statistical probability for outcome O to occur in cases where F is realized is "very high" ($p(O, F)$ is very high) (Hempel 1962, 13-14). For instance, the subsiding of John Doe's hay fever attack (O_i) is explained by the fact that he took eight milligrams of chlor-trimeton (F_i), and the fact that the "probability for subsidence of a hay fever attack upon administration of eight milligrams of chlor-trimeton is high" (Hempel 1962, 13).

Correlations are central to Hempel's I-S model of explanation, and these correlations do not require explanation.²² As Dessler (1991, 344) points out, Hempel's model of explanation is merely the "formalization of...correlational analysis."

²² One might object that some covering laws are, in fact, explanatory. Even though covering laws instantiate a particular relationship as an example of a broader relationship, this broader relationship often invokes processes that exist at a lower level of analysis than the relationship that it explains. This might be the case when, for instance, we explain Galileo's and Kepler's laws as special consequences of Newton's laws of motion and gravity. (Hempel (1962, 11-2) actually argued that this type of relationship is explanatory but not causal.) The problem with covering law explanations, however, is that can fail to distinguish causal from non-causal

In sum, SRs typically assert that mainstream social scientific methods embody a black box approach to causation because they are based on unexplained correlations.

ii. Hume's Philosophy of Causation

For SRs in the social sciences, the roots of the black box approach to causation are found in the penetrating analyses of causation penned by eighteenth century philosopher, David Hume. Indeed, "Humean causation," or Hume's "regularity theory of causation," is synonymous with the black box approach.²³ In this section, I sketch Hume's argument about causation based on his *A Treatise of Human Nature* 1978 [1739] and his *Enquiries Concerning Human Understanding and Concerning the Principles of Morals* 1975 [1748].

Impressions and Ideas

Some background to Hume's general philosophy will provide the foundation on which to build up his analysis of causation.

relationships. Thus, as George and Bennett (2005, 132-3) point out: "...[T]he D-N model allows [a] change in barometric reading to count as an 'explanation' of [a] storm, and cannot distinguish between the explanation via barometric readings and that via air pressure..." In brief, the cover law model of explanation allows in too much.

²³ I do not attempt to trace a direct lineage between Hume and mainstream methodologies for making causal inference, although this lineage could likely be traced. An examination of Hume's "rules by which to judge causes and effects" (1978 [1739], 173-9) might provide evidence of the influence that Hume had on mainstream methodology.

A central pillar of Hume's philosophy, on which he rests numerous arguments, is his analysis of the relationship between *impressions* and *ideas* – the two categories into which “[a]ll perceptions of the mind resolve themselves...” (Hume 1978 [1739], 1). Impressions refer to those perceptions that enter the mind “...with most force and violence...” through a human sense such as sight or touch; ideas, to the “...faint images of these in thinking and reasoning...” (Hume 1978 [1739], 1).

Hume's empiricism leads him to contend that there can be no idea without it being preceded first by a corresponding sense impression: “To give a child an idea of scarlet or orange, of sweet or bitter, I present the objects, or in other words, convey to him these impressions...” (Hume 1978 [1739], 5). Hume notes that “...when one is born blind or deaf...not only the impressions are lost, but also their correspondent ideas...” (Hume 1978 [1739], 5).²⁴

²⁴ Hume (1978 [1739], 5-6) did allow that there are rare exceptions to the assertion that impressions always precede ideas. Hume (1978 [1739], 7-8) also distinguishes a second class of impressions, “impressions of reflexion,” which are preceded by ideas, which are, in turn, preceded by sense impressions: “An impression first strikes upon the senses, and makes us perceive heat or cold...Of this impression there is a copy taken by the mind, which remains after the impression ceases; and this we call an idea. This idea of pleasure or pain, when it returns upon the soul, produces the new impressions of desire and aversion...which may properly be called impressions of reflexion, because derived from it.”

Hume's analysis of the relationship between impressions and ideas is crucial to his analysis of causation, because, for Hume (1978 [1739], 69-75), causation is an *idea*. And because there can be no idea without a corresponding impression that precedes it, Hume's analysis of causation necessarily begins with a search for the *impression* that give rise to the *idea* of causation: "Let us therefore cast our eye on any two objects, which we call cause and effect, and turn them on all sides, in order to find that impression, which produces an idea of such prodigious consequence" (Hume 1978 [1739], 75).

What Necessary Connection Is Not

Hume (1978 [1739], 75-78) identifies three impressions common to all relationships that have the appearance of being causal in nature: (1) "contiguity" – a "cause" is always situated near to or next to its "effect"; (2) "priority of time" or "succession" – a "cause" always occurs before its "effect"; and (3) "necessary connexion" – granted the presence of a "cause," its "effect" must follow.

Although Hume (1978 [1739], 77) finds the first two impressions – contiguity and succession – relatively unproblematic²⁵, he is troubled by necessary connexion (connection). For on closer inspection it becomes plain to Hume that necessary connection is not an impression after all, but rather an *idea*. And again, for Hume there can be no idea without an impression that precedes and corresponds to the idea.

²⁵ Hume (1978 [1739], 232-51) does deal with some problems relating to contiguity; and likewise deals with an objection to succession on p.76.

Hume is especially troubled by his observation that necessary connection is an idea, because he considers necessary connection to be a far more important component of causation than either contiguity or succession: “An object may be contiguous and prior to another, with being consider’d as its cause. There is a necessary connexion to be taken into consideration; and that relation is of much greater importance, than any of the other two above-mention’d” (Hume 1978 [1739], 77).

The next step in Hume’s analysis of causation is therefore to discover the impression that gives rise to the idea of necessary connection: Why do we *believe* that the next time we observe a “cause,” its “effect” will necessarily follow?

In searching, Hume (1978 [1739], 155-172) argues that the “power,” “force,” “efficacy,” or “energy” that binds cause and effect together – what Hume generally referred to as “the cement of the universe” – cannot be experienced by sense impression, and therefore cannot be the source of the impression that gives rise to the idea of necessary connection. In so doing, he rejects outright a variety of arguments from his philosophical predecessors and contemporaries:

There are some, who maintain, that bodies operate by their substantial form; others, by their accidents or qualities; several, by their matter and form; some, by their form and accidents; others, by certain virtues and faculties distinct from all this. All these sentiments again are mix’d and vary’d in a thousand different ways; and form a strong presumption, that none of them have any solidity or evidence... This presumption must encrease upon us, when we consider, that these principles of substantial forms, and accidents, and faculties,

are in reality any of the known properties of bodies but are perfectly unintelligible and inexplicable. (Hume 1978 [1739], 158)

Hume continues the search by rejecting some related avenues from whence the idea of necessary connection might arrive.²⁶ Hume (1978 [1739], 78-84) rejects the related notions that every event must have a cause, and that a particular event must have a particular cause. Hume notes that it is commonly taken for granted "...without any proof given or demanded..." that "whatever begins to exist, must have a cause of existence." Hume rejects this maxim as fallacious, arguing that it is possible to comprehensively define an event without invoking its cause. Moreover, Hume argues that, unlike our inability to conceive of a square without four sides, it is *possible* to conceive of a situation whereby an event is uncaused:

The separation, therefore, of the idea of a cause from that of a beginning of existence, is plainly possible for the imagination; and consequently the actual separation of these objects....implies no contradiction nor absurdity. (Hume 1978 [1739], 79-80)

Hume (1978 [1739], 155-172, 1748, 60-73) also rejects another avenue: that internal, psychological experience might give rise to the idea of necessary connection. In particular, he employs two distinct arguments to dismiss the notion that the movement of a body part is necessarily brought about by volition. First, he generally contends that little is known about the "...secret union of soul and body..." (Hume 1975 [1748], 65).

²⁶ As Hume (1978 [1739], 78) notes, the search initially moves away somewhat from a direct search for "impressions."

More specifically, Hume (1975 [1748], 66-67) asserts that the movement of a body part is more immediately brought about by physiological functions than by volition; and because little is known about the details of this physiology, the notion of a necessary connection become absurd:

Here the mind wills a certain event: Immediately another event, unknown to ourselves, and totally different from the one intended, is produced: This event produces another, equally unknown: Till at last, through a long succession, the desired event is produced. But if the original power were felt, it must be known; since all power is relative to its effect. (Hume 1975 [1748], 66)

Second, Hume argues that if there were a necessary connection between mind and body, it would be the case that, independent of experience, one could learn that willing a body part to move sometimes works (as in the case of willing your leg to move) and sometimes does not (as in the case of willing your liver to move):

We learn the influence of our will from experience alone. And experience only teaches us, how one event constantly follows another; without instructing us in the secret connexion, which binds them together, and renders them inseparable. (Hume 1975 [1748], 66)

What Necessary Connection Is

After rejecting these possible sources of necessary connection, Hume (1978 [1739], 163) asserts that it is not a single instance of cause and effect, but rather a multiplicity of such like instances, that gives rise to the idea of necessary connection:

...[S]uppose we observe several instances, in which the same objects are always conjoin'd together, we immediately conceive a connexion betwixt them, and begin to draw an inference from one to another. This multiplicity of resembling instances, therefore, constitutes the very essence of power or connexion, and is the source, from which the idea of it arises.

“Constant conjunction,” then, is for Hume the source that gives rise to the idea of necessary connection.

Together with contiguity and succession, constant conjunction forms the essence of causation. Hume’s own billiard ball example encapsulates the three characteristics of a causal relationship:

Here is a billiard ball lying on the table, and another moving toward it with rapidity. They strike; the ball which was formerly at rest now acquires a motion. This is as perfect an instance of the relation of cause and effect as any which we know either by sensation or reflection. Let us therefore examine it. It is evident that the two balls touched one another before the motion was communicated, and that there was no interval betwixt the shock and the motion. Contiguity in time and place is therefore a requisite circumstance to the operation of all causes. It is evident, likewise, that the motion which was the cause is prior to the motion which was the effect. Priority in time is, therefore, another requisite circumstance in every cause. But this is not all. Let us try any other ball of the same kind in a like situation, and we shall always find that the impulse of the one produces motion in the other. Here, therefore, is a third circumstance, viz., that of constant conjunction

betwixt the cause and the effect. Every object like the cause produces always some object like the effect. Beyond these three circumstances of contiguity, priority, and constant conjunction I can discover nothing in this cause. (Quoted in Salmon 1998, 193)

But Hume (1978 [1739], 163) stresses that it is not constant conjunction *itself* that produces the idea of necessary connection. Rather, Hume contends that constant conjunction engenders a *psychological trick*. After a certain amount of like instances experienced through the senses, the mind becomes *conditioned to believe* that the future will resemble the past. If sense experience tells us that *As* are regularly followed by *Bs*, the mind becomes conditioned to believe that the next time *A* is experienced, *B* will follow:

The several instances of resembling conjunctions leads us into the notion of power and necessity. These instances are in themselves totally distinct from each other, and have no union but in the mind, which observes them, and collects their ideas. Necessity, then, is the effect of this observation, and is nothing but an internal impression of the mind, or a determination to carry our thoughts from one object to another. (Hume 1978 [1739], 165)

For Hume then, the impression that gives rise to the idea of necessary connection – and hence, causation – is a determination of the mind invoked by constant conjunction.²⁷

²⁷ Hume (1978 [1739], 7-8) refers to this as an “impression of reflexion.”

iii. Conclusions: Looking Ahead

SRs typically assert that, under the lasting influence of David Hume, mainstream social scientists treat causation as nothing more than an unexplained regularity and that mainstream social scientific methods reflect this black box approach to causation. Hume argued that the “cement of the universe” is unknowable: the closest we get to causation is constant conjunction. SRs point out that mainstream methodologies in the social sciences are based on these principles, as they rely almost exclusively on unexplained correlations. There is an element of truth to these SR claims, but Parts II and III of this thesis argue for crucial refinements.

First, I will contend that mainstream social scientists often do, in fact, invoke unobservable entities and processes in order to explain correlations. However, they often do so in the tradition of IE, whereby theoretical explanations are relatively simplistic (because generalizability is emphasized), and whereby the unobservable entities and processes are treated merely as useful fictions. In this sense, SRs may be wrong to assert that causation is “reduced” to correlation in mainstream social science – at least mainstream political science. But in another sense, even IE represents a black box approach to causation, because it does not “take seriously” theoretical explanations: that is, it matters little to IE whether theoretical explanations are accurate or not, so long as they provide “empirical adequacy.”

Second, contra typical SR claims, I do not hold that mainstream methodologies – statistical analysis, Mill’s methods, and Hempel’s models – are *necessarily* synonymous

with a black box approach to causation. SRs are, I believe, correct in their assertion *if* these methods are considered in isolation, or *in and of themselves*.²⁸ But these methodologies do not *necessarily* exclude *non*-black box approaches to causation: in fact, these methodologies are integral to the SSR approach that I develop in this thesis. The distinction between the SSR approach and that of other SR approaches in the social sciences centers, in part, on how these methodologies are employed. I will attempt to show that by the logic of the AfC, these methodologies can usually be employed in a way that makes theoretical explanations fundamental to causal inference.

Before we get to these refinements, however, it is important to get a better understanding of the SR claims. The remainder of Part I explores the general notion that explanations of regularities are integral to a genuine causal analysis.

²⁸ Some philosophers of science and social science do take a black box approach to causation that makes correlation-based methodologies synonymous with causation (e.g., see Glymour et al., 1982, Kenny 1979, and Kincaid 1994).

Chapter III: The Causal Mechanism Approach

In this chapter and Chapter IV, I review two approaches to causation based in scientific realism: the causal mechanism approach and the causal power approach. Both SR approaches have been developed in response to the Humean-based black box approach to causation outlined in Chapter II.

These two SR-based approaches argue that it is crucial to determine the unobservable entities and processes that explain correlations for philosophical, methodological, pragmatic, and purely intellectual reasons. Philosophically, if causation is going to be a genuinely meaningful concept, then it cannot be reduced to correlation.

Methodologically, explanatory processes can help identify spurious relationships.

Pragmatically, human intervention aimed at either fostering or preventing causal relationships becomes more possible when explanatory processes are known. Finally, intellectually, there is a deep-seated desire in humans to know why things function the way they do.

The philosophers and methodologists who champion the causal mechanism and causal power approaches to causation are not a cohesive group in any way. First, they hail from a variety of academic disciplines in both the natural and social sciences. Second, their works only sometimes engage each other – i.e., they generally do not form an intellectual community. Third, their approach to causation is based on several different philosophical foundations. Some are self-styled *scientific realists*, others are *critical realists*, and still

others admit to no particular philosophical affiliation. What does bind proponents of these scientific realist-based approaches, however, is a belief that Hume's "cement of the universe" is "knowable" and causation is a meaningful concept.

Eschewing the black box approach to causation does not necessarily make one a SR.

There are scholars that reject the black box approach and its related methodologies on other philosophical bases – postmodernists, for instance. Chapters III and IV review only those scholars that *explicitly* defend the causal mechanism and/or causal power approaches, or those that *implicitly* defend these approaches based on their language and ideas.

Much of the SR literature treats causal mechanisms and causal powers as one and the same. I contend, however, that the two are analytically distinct, and I therefore treat them accordingly. In this chapter I review the causal mechanism approach before turning to the causal power approach in Chapter IV.

In the first section of this chapter, I review the arguments of some causal mechanism proponents in the natural sciences, focusing on the influential work of Wesley Salmon.

In the second section, I review some arguments for the causal mechanism approach adopted by social scientists. This literature review is not meant to be comprehensive.

Instead, I focus on what I consider to be the most notable attempts to establish the causal mechanism approach in the natural and social sciences. In the conclusions, I point to

some problems with the methodological remedies championed by causal mechanism proponents, and suggest how Part II of this thesis might help remedy these problems.

i. Causal Mechanism in the Natural Sciences

Proponents of the causal mechanism approach to causation insist that mechanisms are integral to causation. Whereas mainstream causal analysis argues that *causation* = *regularity*, proponents of the causal mechanism approach argue that *causation* = *regularity* + *mechanism*. A mechanism, in the literature, typically refers to the unobservable processes that lies between cause and effect, and explains the relationship. (A more detailed discussion of the definition of mechanism is taken up in Chapter VI.)

Harré (1985, 39-40) provides a concise example that illustrates the point about causal mechanism. The constant conjunction of sun and water with plant growth provides a tentative law but does not explain *why* sun and water make plants grow. To understand why, we need to understand the *mechanism* of photosynthesis. Thus, Harré concludes, “In practice we never rest content with laws for which there are no explanations.”

Recent research into the link between smoking and lung cancer provides another case in point. Although the correlation between smoking and cancer has been known for many years, only within the last decade have researchers pinpointed exactly *why* smoking engenders cancer. That is, researchers have pinpointed the causal mechanisms that link the inhalation of smoke to the production of cancerous cells in the body (Grady 1996, 3, Thagard 1999, 109). Similarly, although scientists have known for nearly a century that

aspirin relieves pain, it is only within the last decade that they have discovered the causal mechanisms behind this pain relief (Garavito 1999, 108).

Taking up this line of argumentation, Thagard (1999, 37-70) uses recent medical research to show that the identification of causal mechanisms has played an important role in solidifying the causal relationship between the bacteria, *Helicobacter Pylori*, and ulcers.²⁹

Thagard (238-240) argues that the success of recent research into the relationship between *Helicobacter Pylori* and ulcers can only be understood through the lens of SR.

According to Thagard (101), the bacterial theory of ulcers originated in 1982 when two Australian physicians, Barry Marshall and J. Robin Warren, noticed an association between ulcers and infection with *Helicobacter Pylori*, a previously unknown bacterium that Warren had microscopically discovered in biopsy specimens in 1979. Although Marshall and Warren's epidemiological study found a significant correlation between *Helicobacter Pylori* and ulcers, the medical community largely dismissed their claim because there was *no known mechanism* by which bacteria could cause an ulcer. More specifically, it was believed that the stomach contained too much acidity to tolerate the presence of bacteria. Moreover, as Thagard (113) points out, "correlation themselves

²⁹ Thagard actually endorses a causal power approach in addition to a causal mechanism approach (102-104). But in my estimation, his analysis fits better with the latter than the former. This is an example of how these two approaches are conflated in the literature. Thagard argues that his approach is compatible with that of Wesley Salmon (116) and Nancy Cartwright (103), even though, as we shall see, Salmon and Cartwright take decidedly different approaches.

have no explanatory force, since they may be the result of confounding alternative causes.”³⁰

The prevailing theory, linking ulcers to excessive acidity, held the day until medical researchers in the 1990s discovered *Helicobacter Pylori* produce ammonia that neutralizes stomach acid. That is, the prevailing wisdom connecting ulcers to acidity could not be displaced until the mechanism linking *Helicobacter Pylori* and ulcers was discovered. When mechanisms “are known or plausible,” concludes Thagard (112), “...they can enhance the explanatory coherence of a causal hypothesis. Moreover, causal hypotheses incompatible with a known mechanism are greatly reduced in explanatory coherence.”

The Physical Approach to Causal Mechanism

As the examples above demonstrate, the argument for causal mechanism in the philosophy of the natural sciences generally involves a search for the unobservable biological and physiological processes linking causes and effects. Elsewhere in the philosophy of natural science, the search for causal mechanisms has taken on a somewhat different form. A small group of philosophers of science have developed a theory of *physical causation* as an alternative to a Humean-based black box approach. Wesley

³⁰ It is worth noting that in a private e-mail communication on May 31st, 2001, Thagard makes this claim even for ‘true’ experiments because “it is never possible to control for *all* possible experimental confounds.”

Salmon (1984, 1998), has led the way in developing this physical approach to causation. In what follows I will explicate the arguments made by Salmon.

Salmon's work is worth investigating in detail for at least three reasons. First, even after his untimely death in 2001, Salmon continues to be regarded as a leading philosopher of science.³¹ Second, I will make extensive use of Salmon's work in Chapter IX to demonstrate a point relating to causation.

Third, and most poignantly, Salmon himself was once a strong proponent of a black box approach, as illustrated in his 1971 *Statistical Explanation and Statistical Relevance*. Following Hempel (1962, 9-33), most black box causation proponents had always maintained that an event could only be explained by a 'statistical law' that inferred a 'high' probability on that event.³² Salmon, however, argued that in order for a factor to qualify as a cause of a given outcome, the presence of that factor need only *change* the probability of the outcome *relative to* the probability of the outcome in the absence of the factor, or $P(E/C) \neq P(E/\sim C)$. Salmon called his causal account the Statistical-Relevance or S-R model of scientific explanation, in contrast to Hempel's I-S model of explanation. Thus, while most proponents of black box causation demanded a 'high-probability'

³¹ In addition to a long list of publications, Salmon boasts an entry devoted to his philosophy in the *Oxford Companion to Philosophy* (Honderich 1995, 787-788). He also has three books dedicated to his honour (Fetzer 1988, Galavotti and Pagnini 1999, and McLaughlin 1982).

³² Although Hempel never clearly defined what he meant by "high probability," some believe that he meant over fifty percent, and probably closer to seventy or eighty percent.

relationship, Salmon argued that ‘low-probability’ relationships – and even ‘negative-probability’ relationships³³ – were causal. In short, Salmon was a Humean *par excellence*.

Not long after developing his S-R model of scientific explanation in the early 1970s, Salmon’s views on causation began to change.³⁴ By the late 1970s, Salmon had adopted a radically new view of causation, which he propounded in his 1984 *Scientific Explanation and the Causal Structure of the World*:

In [my previous work], I advocated what came to be called the statistical-relevance or S-R model of scientific explanation. At that time, I thought that anything that satisfied the conditions that define that model would qualify as a legitimate scientific explanation. I no longer hold that view. It now seems to me that the statistical relationships specified in the S-R model constitute the *statistical basis* for a bona fide scientific explanation, but that this basis must be supplemented by certain *causal factors* in order to constitute a satisfactory scientific explanation. (Salmon 1984, 34)

His new view on causation, Salmon (1984, xi) argues, would require “nothing less than a radical gestalt switch in our thought about scientific explanation.”

³³ In making this argument, Salmon lowered the bar even further than did Suppes (1970), because Salmon’s analysis admitted even those causes that *lowered* the probability of their effect.

³⁴ E.g., see Salmon (1977).

What engendered this sea change in Salmon's views? The answer to this question lies in his deep-seated dissatisfaction with Humean causation: an impassioned desire to "put the *cause* back into *because*" (Salmon 1998, 193). By the standards of mainstream causal analysis, Salmon (1984, 4-5) argues, scientific explanation allows only for prediction. He points out that Hume and mainstream causal analysts are deeply skeptical about causation noting that although Hempel suggested he was dealing with *causal* explanation when he first introduced his Deductive-Nomological (D-N) model of explanation in 1948, he later backed away from this claim (Salmon 1998, 194).

Salmon's disenchantment with "non-causal explanation" led him to develop a physical approach to causation that has taken on two successive versions; the first defended in his 1984 *Scientific Explanation and the Causal Structure of the World*, and the second in his 1998 *Causality and Explanation*. Salmon (1998, 204) claims that both versions provide "a solution to Hume's problem of causality" by providing an account of causation that does not rely on constantly conjoined events, but relies instead on the fluid process of causal mechanism.

Salmon's first version of a physical approach to causation – which I will call his *mark transmission theory of causation* – is concerned with differentiating causal relationships from non-causal relationships. According to Salmon, the key to making this distinction lies in discovering the true nature of mechanisms.

The mark transmission theory of causation can be illustrated with an example offered by Salmon. Imagine a rotating spotlight, mounted in the center of a circular room, which casts a spot of light on the perimeter wall. The spot of light on the wall moves in a circular motion around the wall as the spotlight rotates. Given this setup we can observe two processes. The first is the transmission of the light beam from the spotlight to the wall. The second is the spot of light moving in a circular motion around the circumference of the wall. But only the first of these two processes are actually *causal*. The first process is causal because it is the light beam that causes the wall to illuminate: The light beam that emanates from the spotlight has an internal structure that is propagated from the spotlight to the wall. The second process is *not* causal because the spot of light at one point on the wall does not *cause* the continuous series of spots of light as it rotates around the room. The spot of light on the wall carries an internal structure but does not propagate this internal structure around the wall. In Salmon's terms, the first process is "genuinely causal" while the second is merely a "pseudo-process."

But how do we differentiate the genuine process from the pseudo-process? How do we know that one process is propagating its internal structure while the other is not? We verify by inducing a modification in the internal structure of the process – what Salmon calls a "mark" – and check to see if this mark is transmitted. In this example, one way to induce a modification in internal structure is to place a red filter in the path of both observed processes. If we placed a red filter anywhere along the path from the spotlight to the wall, this mark would be transmitted to the wall where the spot of light would now assume the colour red. That is, the modification in internal structure would be

transmitted. If we placed a red filter along the wall, however, the spot of light passing through this filter would only *momentarily* assume the colour red. Once the spot of light had passed through the filter it would once again assume its original colour (white). In this second situation, the mark – the modification in internal structure – would *not* be transmitted, alerting us to the fact that this second regularity was merely a pseudo-process.

Consider a second example to illustrate Salmon's theory. A car travels along the highway on a sunny day. As the car moves, its shadow moves in parallel motion along the shoulder of the highway. The moving car is a genuine causal process, its internal structure being propagated as it hurtles down the highway. The moving shadow is a mere pseudo-process; although its internal structure may *appear* to be propagated, it is just an illusion. If the car collides with a brick wall, it will carry the 'marks' of that collision – the dents and scratches – long after the collision has taken place. If, however, only the shadow of the car 'collides' with the brick wall, it will be deformed momentarily, but it will resume its normal shape just as soon as it passes beyond the wall. Again, only causal processes propagate internal structure and transmit marks; non-causal pseudo-processes do neither.

With this first version of physical causation, Salmon (1998, 191, 248) claimed to "answer Hume's basic challenges to the concept of causality. Causal processes, as characterized by this theory, constitute precisely the objective physical causal connections which Hume sought in vain."

The second version of physical causation – which I will call his *conserved-quantities theory* – focuses on transmission of energy rather than transmission of a modification to internal structure. In addition, this second theory invokes the concepts of “conserved quantities” and “causal interaction.” A causal interaction involves a transmission of conserved quantities, which represents a mechanism. For instance, in an interaction involving an exchange of momentum, the total momentum of the outgoing processes must be roughly equal to that of the incoming processes. This transmission of energy represents a causal mechanism.

Again, an example provided by Salmon is useful to illustrate this second physical version of causation. Consider a baseball and a molecule of nitrogen that strike a glass window simultaneously, causing the window to shatter. How do we know that it was the baseball and not the nitrogen molecule that caused the window to shatter? We could use a counterfactual approach and reason that the window would not have broken had it not been hit by the baseball. But Salmon insists that we need not appeal to counterfactuals. Instead we need only rely on the ideas of causal interaction, and transmission of conserved quantities. Quite simply, the momentum of the incoming nitrogen molecule would be drastically smaller than the momentum of the outgoing shards of glass as they fanned out into the air. The momentum of the incoming baseball, conversely, would be roughly equal to that of the outgoing shards of glass. The causal interaction of the baseball and the window involves a transmission of conserved quantities; the causal interaction of the nitrogen molecule and the window do not.

Salmon (1998, 204) argues that both his theories provide “a solution to Hume’s problem of causality.” “Whether causal processes are characterized in terms of mark transmission or transmission of conserved quantities,” proclaims Salmon, “the notion that causal processes provide physical causal connections has deep philosophical import...The importance of the distinction between this physical approach and the approach that analyzes causality in terms of constant conjunction and/or statistical correlations cannot be overemphasized” (Salmon 1998, 200).³⁵

ii. Causal Mechanism in the Social Sciences

The search for a theory of causation that “opens up” Hume’s black box has also consumed some philosophers of social science. In fact, in the social sciences, the need to think in terms of causal mechanism is sometimes seen as perhaps even more important than in the natural sciences. As noted above, one of the factors that motivate SRs is the need to overcome the potential for spuriousness. In the social sciences, this potential is more acute than in the natural sciences, because, whereas it is often possible to conduct laboratory or randomized experiments in the natural sciences, the social sciences are

³⁵ As Salmon readily acknowledges, Phil Dowe originated the ‘conserved quantity’ theory of causation, in response to what Dowe saw as problems with Salmon’s ‘mark transmission’ theory. Likewise, Dowe acknowledges his intellectual debt to David Fair (1979), who in turn derived his own theory of physical causation from Jerrold Aronson (1971). In all accounts, causation can be understood as a transmission of energy from one body to another. And in all accounts, the authors direct their theories of physical causation explicitly against Hume.

often confined to quasi-experimental techniques such as statistical analysis or Mill's methods.³⁶ Quasi-experimental methods are widely considered inferior to experimental methods because they are less able to guard against spuriousness.³⁷

Daniel Little is one social scientist that embraces a causal mechanism approach, in his 1991 *Varieties of Social Explanation*.³⁸ Little (1991, 14) defines his "causal mechanism thesis" as: "C is a cause of E [if] there is a series of events C_i leading from C to E, and the transition from each C_i to C_{i+1} is governed by one or more laws L_i ." He contrasts this causal mechanism thesis with the Humean "inductive regularity thesis," which he defines as: "C is a cause of E [if] there is a regular association between C-type events and E-type events."

To support his argument that causal mechanisms are needed to show *whether* a regularity is non-spurious, Little (1991, 174) invokes the familiar disjuncture between correlation and causation:

³⁶ Of course, many natural sciences – such as epidemiology, evolutionary biology, and ecology – are also stuck, at least in part, with quasi-experimental methods.

³⁷ The difference between experimental and quasi-experimental methods is one of degree not kind: even 'true' experiments are vulnerable to spuriousness. Moreover, as noted above, SRs are not solely driven by methodological concerns, but also by philosophical, pragmatic, and intellectual concerns.

³⁸ Although Little (1991) does not explicitly claim that his argument is underpinned by scientific realism in this book, he does so in his 1998, *Microfoundations, Method, and Causation*. This latter book is discussed in Chapter IV because it adopts a causal power approach.

...A statistical study can provide empirical grounds for accepting or rejecting a causal hypothesis, but the statistical findings themselves are not final or conclusive. Study of covariance among factors is a useful tool for investigating causal hypotheses, but it is always possible that the causal hypothesis is true although the corresponding statistical test is negative.

To counteract the ever-present possibility of spuriousness, contends Little (1991, 178), we “...must be able to identify, at least approximately, the causal mechanisms that underlie the statistical regularity.”

But Little (1991, 178) notes that the case for causal mechanisms does not merely comprise the need to know *whether*, but also the desire to know *why*:

There is a deeper consideration as well that militates against rock-bottom statistical explanations... The demand for an explanation of an event or regularity typically involves this question: Why did this event come about, given the circumstances at the time of occurrence? This is a demand for a causal story, which requires an account of the laws and mechanisms through which antecedent conditions brought about the explanandum.

Given their limitations, Little argues that statistical explanations cannot stand alone: “The discovery of a statistical regularity among variables,” contends Little (1991, 179), “constitutes an empirical description of social phenomena that itself demands explanation.”

Little's overarching conclusion is that causal mechanisms are *fundamental*; indeed, more fundamental than constant conjunction. "I therefore conclude," writes Little (1991, 25), "that the inductive regularity criterion is secondary to the causal mechanism criterion: There is a causal relation between two variables if and only if there is a causal mechanism connecting them."

Andrew Bennett and Alexander George likewise embrace the notion of causal mechanism.³⁹ Bennett and George base their argument on a distinction they draw between "causal effect" and "causal mechanism." Bennett (1997, 7-8) defines causal effect as "the change in probability and/or value of the dependent variable that would have occurred if the explanatory variable had assumed a different value." Causal mechanism, on the other hand, is defined as "the causal process and intervening variables through which causal or explanatory variables produce causal effects." "The difference between a law and a mechanism," contend George and Bennett (2005, 141), "is that between a static correlation ('if X, then Y') and a 'process' ('X leads to Y through steps A, B, C')."

³⁹ The work of Bennett and George comes from a series of unpublished papers (some authored only by Bennett) that are, in part, incorporated in George and Bennett (2005). In what follows, I will cite from the specific unpublished papers from which the ideas are derived, as well as from George and Bennett (2005).

Bennett (1997, 8) and George and Bennett (2005, 137, 140) argue that Hume's own analysis of causation provides a rationale for thinking in terms of causal mechanism. They contend that of Hume's three criteria of causation – constant conjunction, contiguity, and temporal succession – the focus of attention has been primarily on the first, constant conjunction.⁴⁰ But constant conjunction, argue Bennett and George, relates only to causal effect, whereas contiguity and temporal succession relate to causal mechanism. Causal effect, they warn, is *not* synonymous with the definition of causality and we should not relegate causal mechanism to an “inferior status.” Neither causal effect nor causal mechanism is more fundamental in any way; neither is logically or ontologically prior to the other, and both are necessary to a comprehensive definition of causality (Bennett 1997, 9-10).

According to political scientist David Dessler (1991), an emphasis on causal mechanism is needed to move the theory of war beyond evidence based on mere correlation to a genuinely causal theory of war.⁴¹ Dessler (337-338) reviews the results of the Correlates of War (COW) project – “the most systematic and extensive effort in the history of the

⁴⁰ As Chapter II's discussion of Hume should make clear, George and Bennett are not entirely accurate to suggest that, for Hume, constant conjunction, contiguity, and temporal succession, are three equal criteria of causation. In fact, the latter two criteria were far less important to Hume than constant conjunction.

⁴¹ Later, we will see that Dessler also appeals to the notion of causal powers. For now, however, I will examine his appeal to causal mechanisms.

study of war” – probing the reasons for its failure to produce a theory of war.⁴² “From the earliest years,” notes Dessler, the COW’s stated goal was “an integrated and coherent theory of war. But both inside and outside the project, reviewers today acknowledge that the long-sought crystallization of our fragmented knowledge remains an unfulfilled hope.” What accounts for this failure? According to Dessler, the answer lies in an overemphasis on correlational evidence combined with a clear disregard for evidence based on causal mechanism.

Finally, noted philosopher of social science, Jon Elster, also endorses a causal mechanism approach.⁴³ In Elster’s 1983, *Explaining Technical Change*, he contends: “Understanding the details of the causal story reduces the risk of spurious explanations (i.e., mistaking correlation for causation). Also, knowing the fine grain is intrinsically more satisfactory for the mind.”⁴⁴ For Elster (1983, 24), mechanisms are necessary for a complete

⁴² A number of findings have been generated by the COW project, including the democratic peace hypothesis. See Chapter XII for more details on the genesis of the democratic peace hypothesis.

⁴³ As we will see, in Elster’s three main publications on the issue of causal mechanism, there is a self-acknowledged transition in his approach to the issue – from the articulation of a causal mechanism approach to the articulation of a causal power approach. I will examine Elster’s defence of the causal mechanism approach here before examining his causal power approach below.

⁴⁴ This quotation actually comes from Elster’s later work, (1998, 4), but he uses it to encapsulate his position in his 1983, *Explaining Technical Change*. See especially Elster (1983, 26) where he argues that laws of association may always be “epiphenomenal” (i.e. spurious) and therefore cannot be considered causal.

explanation: “To explain is to provide a mechanism, to open up the black box and show the nuts and bolts, the cogs and wheels of the internal machinery.” Mechanisms close the spatio-temporal gap between explanans and explanandum by providing a “continuous and contiguous chain of causal...links.”

iii. Conclusions: Pragmatic Implications

The SRs reviewed in this chapter dispute the Humean-based black box approach to causation, arguing that mechanisms are at least as important as regularities, if not more so. But, except perhaps for the Salmon’s physical causation, I have not yet discussed the methodological implications of the causal mechanism approach.

How exactly do causal mechanism proponents propose that mechanisms be identified? If they do offer advice on this front, most causal mechanism proponents in the social sciences argue that either case studies (e.g., see Bennett 1997, Dessler 1991, George and Bennett 2005, Little 1991) and/or rational choice modeling (e.g., see Little 1991, 1998) should be used to identify explanatory mechanisms.

I agree with causal mechanism proponents on this point, but there are serious shortcomings in these arguments that I attempt to rectify in Part II of this thesis. Most importantly, SRs that advocate a causal mechanism approach sometimes assume that the mechanisms identified by case-study methods and rational choice modeling can stand on their own without necessarily deriving and testing their observable implications. I contend, conversely, that although these methods provide partial tests of the mechanisms

they identify, further testing is required against correlational (and non-correlational) empirical phenomena.⁴⁵ In Chapter X, I argue that this is a key basis of the argument from coincidence (AfC), a core component of SR.

Before getting to these modifications, however, I turn to the second SR-based approach to causation: the causal power approach.

⁴⁵ My thesis deals primarily with a particular type of case-study methodology: namely, process-tracing. But my argument applies equally to other case-study methods, such as comparative case-study methods: observable implications derived from mechanisms identified by these methods should be tested against empirical reality.

Chapter IV: The Causal Power Approach

Some SRs base their approach around the concept of causal power rather than causal mechanism. Causal power refers literally to the *power* of an entity to cause something. For instance, electric fans have the power to circulate air, calculators have the power to compute numbers, and volcanoes have the power to erupt. As we will see, in lieu of the term ‘power’, some scientific realists prefer the terms, *capacity*, *tendency*, or *propensity*.

With the causal mechanism approach, the goal is to discover the process by which one event causes another. With the causal power approach, the goal is to determine the *internal structure* that gives rise to a causal power. For instance, it is certain geological, chemical, and physical properties that constitute the internal structure of a volcano and give it the power to erupt. Likewise, it is the mechanical and physical structure of fans that give them the power to circulate air. In the causal power approach, mechanism does play a role, but only insofar as it represents the *activation* of a causal power. Volcanoes have the *power* to erupt, but this does not mean they are constantly erupting. In most cases these powers will lie dormant for years or even millennia until a geological event activates the volcano, at which point the power becomes a mechanism.

The first section of this chapter reviews the causal power approach in the natural sciences. The second section reviews this approach in the social sciences. Finally, in the conclusions I argue that, even if causal power proponents fail to draw out any

methodological implications from their approach, this issue is moot: more troublesome are philosophical problems with the causal power approach that I point to in Chapter V.

i. Causal Power in the Natural Sciences

One of the leading philosophers of science to advocate the causal power approach is Nancy Cartwright. Most notably, Cartwright makes the case for this approach in her 1989, *Nature's Capacities and Their Measurement*.

Cartwright (3) begins her presentation of the case for causal powers by pitting her position against that of Hume. She notes that for Hume, causation is not genuinely meaningful because “generic [i.e. general] causal facts are reducible to regularities.” Instead of seeing causes as mere regularities, she argues that we should see causes as capacities⁴⁶:

The generic causal claims of science are not reports of regularities but rather ascriptions of capacities, capacities to make things happen... ‘Aspirins relieve headaches.’ This does not say that aspirins always relieve headaches, or always do so if the rest of the world is arranged in a particularly felicitous way, or that they relieve headaches most of the time, or more often than not. Rather it says that aspirins have the capacity to relieve headaches, a relatively enduring and stable capacity that they carry with them from situation to situation; a capacity which may if circumstances are right reveal itself by producing a regularity...

⁴⁶ Although I assume that Cartwright uses “capacities” synonymously with “causal powers,” she would likely deny that capacities have all the attributes that other causal power proponents associate with causal powers.

In addition to countering Hume on this point, Cartwright (2-3) also opposes Hume's dismissal of the single case. Recall that for Hume, a single instance of a putative causal connection is insufficient. Instead one has to observe a regular succession of like events before making any claims about the single instance. Cartwright notes that, "For Hume, singular causal facts are true in virtue of generic causal facts." Conversely, Cartwright contends that the single instance is primary because "the methods that test causal laws by looking for regularities will not work unless some singular causal information is filled in first." Moreover, regularities are merely "evidence...that certain kinds of singular causal fact have happened." For Cartwright "one good single case" can reveal a causal capacity. "The best sign that aspirins relieve headaches," she argues, "is that on occasion some of them do." Thus, Cartwright notes that her position is "doubly anti-Humean" because she endorses capacities, not regularities, and because she takes the single case to be more fundamental than the generic one.

A closer look at Cartwright's arguments will help make sense of these bold claims. An appropriate starting point is Cartwright's assertion that regularities can identify causal relationships, but *only because they presuppose certain information that is already causal*. Hume argued that causal relationships boil down to mere regularities.

Cartwright counters that because regularities presuppose causal information, Hume had it backwards (179-182).

Cartwright begins her defence of this argument by providing what can be considered a universal Humean criterion for establishing the symmetry between correlation and causation. It is well recognized that to test for a causal relationship between a putative cause C and an effect E , it is not enough to compare $P(E/C)$ with the $P(E/\sim C)$. This is because even if $P(E/C) > P(E/\sim C)$, the regularity may be due to a third cause – i.e. the relationship between E and C may be spurious. Instead, we must compare $P(E/C)$ and $P(E/\sim C)$, with all other factors (F_1, \dots, F_n) held constant. This is the criterion that most modern day Humeans⁴⁷ would agree turns correlation into causation.⁴⁸ But as Cartwright (55-56) notes, the conditioning factors, F_1, \dots, F_n , must include every single factor, other than C itself, that either *causes* or *prevents* E . In other words, causation that is based on regularities *presupposes* causal information. “Followers of Hume would like to reduce causal claims to facts about association,” argues Cartwright, but the criterion for doing this is “incomplete until the conditioning factors are specified.” How do we specify the factors to hold fixed? The answer is simple: we identify them as causes. As a result, concludes Cartwright (57), this widely accepted criterion cannot “provide a way to replace the concept of causality by pure association.” This is why Cartwright champions the slogan, “No causes in, no causes out.”

⁴⁷ As will be seen in Chapter X, most modern-day Humeans also believe in unobservable regularities.

⁴⁸ Kincaid (1994, 117), for instance, argues “To confirm that A causes C , I must find a significant correlation between them while controlling for other factors B_1, \dots, B_n which might be the real cause of C ...”

Recall that Cartwright (89) claims Hume had it “upside down” not only because he endorsed regularities, but also because he failed to see the importance of the single case. Cartwright argues that not only is prior causal knowledge required in order to turn correlation into causation, but the causal knowledge that *is* required can be found at the *singular* rather than the general level.

To understand why this is the case, recall that according to the widely accepted criterion for turning correlation into causation, $P(E/C) > P(E/\sim C)$ provided that F_1, \dots, F_n are held constant. But Cartwright notes that this criterion is actually too strong. It holds fixed factors that *follow* the putative cause, ‘C’, whereas what we really want to do is hold fixed only those factors that are *prior* to the putative cause ‘C’. The reason for this is simple. If we hold fixed the intermediary factors between ‘C’ and ‘E’, ‘C’ might falsely appear to have no effect, or at least falsely appear to have a reduced effect.⁴⁹ It is not that Humeans fail to recognize this methodological lesson, but rather that if they apply this in practice, they are at a loss to make sense of certain causal arrangements, such as the one depicted in Figure 1.

In this causal arrangement, ‘E’ is produced by ‘C’ via ‘F’, but ‘F’ can also occur on its own (or perhaps sparked by some other factor, say ‘G’) to produce ‘E’. Given this arrangement, the Humean faces a conundrum. If you control for ‘F’, you risk reducing or annulling the actual effect of ‘C’. But if you fail to control for ‘F’, you may fail to detect

⁴⁹ In Chapter XIII, I dispute this methodological rule, arguing that controlling for factors after the main IV can help isolate mechanisms.

that the relationship between 'C' and 'E' is spurious. The solution to this dilemma, argues Cartwright, is to separate those instances in which 'C' engenders 'F' from those in which 'F' occurs independently or via some any other factor. The subpopulation that consists solely of those instances in which 'F' is "causally influenced" by 'C', then, provides the test population by which to judge the causal relationship between 'C' and 'E'. But the dilemma for the Humean is now twofold: "This means that what counts as the right population in which to test causal laws by probabilities will depend not only on what other causal laws are true, but on what singular causal processes obtain as well. One must know, in each individual where 'F' occurs, whether its occurrence was produced by 'C', or whether it came about in some other way" (Cartwright 1989, 96). Cartwright concludes that the criterion for turning correlation into causation has to be amended to include this type of further subdivision of population; but to make this amendment is to admit the need for prior causal knowledge at the singular (as opposed to general) level.

Cartwright (97) provides an example to illustrate the causal arrangement in Figure 1 and the amended criterion for turning correlation into causation. Imagine a situation whereby on some (but not all) Monday afternoons, you call me at my house on the west coast from your house on the east coast. Here, 'C' (your dialing my number) acts through 'F' (my phone ringing) to cause 'E' (my picking up the phone). According to the accepted criterion for turning correlation into causation, we would control for all factors that precede 'C', in order to determine if 'C' is a cause of 'E'. Assume that this is the case. But we *do not* control for factors that follow 'C', because this would falsely

eliminate or reduce the impact of the causal process initiated by 'C'. Therefore we do not control for 'F', my phone ringing. That is, we do not look within the 'F' and the '~F' populations separately. If we *did* control for 'F', we would find that in both the 'F' and '~F' populations, $P(E/C) = P(E/\sim C)$. In the F population, $P(E/C) = 1$ and $P(E/\sim C) = 1$. In the '~F' population, $P(E/C) = 0$ and the $P(E/\sim C) = 0$. So, Cartwright (97) concludes: "Holding fixed 'F' in this case would give a misleading causal picture."

So, the solution would appear to simply allow 'F' to vary naturally – i.e. do not hold 'F' fixed, or do not look within the 'F' and '~F' populations separately. But doing this can also provide a misleading picture. Imagine the same example, but on each Monday afternoon another friend of mine, just a little closer, calls me at the same time, and you never succeed in getting through to me. In this case, if we don't control 'F', it looks as if 'C' (your dialing my number) is actually causing 'E' (my picking up the phone), since now $P(E/C) > P(E/\sim C)$. But we know that this is not the case because it is the *other* friend's calling that is causing 'E' (my picking up the phone) through 'F' (my phone ringing).

Cartwright's example appears to illustrate the dilemma at hand – both controlling for 'F' and allowing it to vary causes problems given the causal arrangement depicted in Figure 1. According to Cartwright's solution to this conundrum, we need to parse the populations. First, we need to look at the population that includes cases in which 'F' is present but only because it is caused by 'C'. We can call this population '~F*'. Second, we need to look into the population that excludes cases in which 'F' is present but is

caused by 'C'. We can call this population 'F*'. If we made this parsing, we would find that in ' $\sim F^*$ ', $P(E/C) > P(E/\sim C)$. The exact probability would depend on how many of those Monday afternoons my phone rang because of your calling and how many days it did not ring, but we can be sure given this causal arrangement that $P(E/C)$ would be larger than $P(E/\sim C)$. This makes sense since ' $\sim F^*$ ' includes only those Monday afternoons on which my phone would not otherwise ring (were it not for your calling me). In 'F*', we would find that $P(E/C) = P(E/\sim C)$. Again the exact probabilities would depend on the distribution of days in which the phone rang versus days in which it did not ring. Regardless, the equality would be assured, which also makes sense since 'F*' includes only those Monday afternoons on which my phone rings only because my other friend has called. With this solution then, the probabilities properly reflect the causal arrangement.

Cartwright's solution to the conundrum seems to work. So, what is the problem?

Cartwright's claim is that her solution admits too much for the Humean. First, we need *prior* causal knowledge at the general level ("What factors, 'F', should we hold fixed?"). And second, we need causal knowledge at the singular level ("Is this one of the cases of 'F' that should be put into 'F*' or ' $\sim F^*$ '"). This is why Cartwright argues that Hume had it backwards. In order to properly assess causal relations, we cannot simply rely on regularities. We need prior causal information that precedes the assessment of these

regularities. Moreover, this prior causal information we need must, at least sometimes, come at the singular level – not at the general level at which regularities operate.⁵⁰

So far, I have summarized Cartwright's argument against Hume. But how does this play into her endorsement of causal powers, or as she prefers, capacities? To understand this connection, return for a moment to the telephone example. In this example, there is a causal law connecting 'C' and 'E', but that law only holds in the population ' $\sim F^*$ '. For Cartwright, this shows that *causal laws are relativized to particular populations*. That is, laws are not general – i.e. they are not invariant over every possible arrangement of 'E's' other causes. But Cartwright claims that, given the truth of this proposition, the important assumption of *contextual unanimity* is breached. According to Cartwright, in standard Humean probabilistic accounts of causality, a genuine cause must increase the probability of the effect in *every causally homogeneous population*. That is the requirement of *contextual unanimity*. What laws describe, however, is only a *limited contextual unanimity*, namely, 'contextual unanimity' *within a circumscribed population*. So, it turns out that we need another concept *at a higher level of generality* than 'mere' laws, to account for causal relationships that are invariant over all background conditions – i.e. a concept that allows for genuine 'contextual unanimity' (142-145).

⁵⁰ I believe that Cartwright's claim could even be strengthened. She fails to point out that, given her solution in parsing the populations into an F^* and $\sim F^*$, you would also need some prior knowledge to know in which of these two subpopulations to place the days that the phone does not ring at all.

The final step of Cartwright's argument is to contend that the concept of 'capacities' is just the "concept of causation that philosophers have been grasping for in their various probabilistic theories of causality" (Cartwright 1989, 145). Capacities operate at a *higher* level of generality (or, as she prefers, "modality") than do causal laws. As noted above a capacity refers to the causal power of an entity. This capacity is something that an entity carries with it in *under all circumstances and background conditions*. Aspirins have the capacity to relieve headaches regardless of who takes the aspirin and under what conditions the aspirin is taken. However, a capacity may not always reveal itself in the probabilities. This is not because the entity has lost its capacity, but because in that particular population – i.e., in that particular arrangement of causes – there *may be other capacities that override or distort that particular capacity*. Aspirins have the capacity to relieve headaches, a capacity that they carry with them "from situation to situation." But under certain circumstances this capacity may not reveal itself, maybe because the person taking it has developed an immunity to its effects, or maybe because the aspirin is taken under an atmospheric pressure that annuls its impact. Who knows? But this does not mean that the aspirin no longer has the *capacity* to relieve headaches. It only means that there are other capacities at work in this particular population that interfere with its impact and make it look – in the probabilities – as if aspirins do not relieve headaches. But probabilities can be deceptive Cartwright maintains, because probabilities only reveal laws that are limited in terms of contextual unanimity. True contextual unanimity can only be found in the concept of capacity. Thus Cartwright (1989, 145) sums up: "To believe in contextual unanimity is to believe in capacities."

I can best sum up Cartwright's argument as follows. The Humean conducts a statistical analysis and finds a constant conjunction. This constant conjunction may be deterministic, $P(E/C) = 1$, or probabilistic, $P(E/C) < 1$. This leads them to declare a law that is reflected in the regularity. But the regularity is then scrutinized under different background conditions and it is found that it no longer holds. Does this mean that there is no causal law after all? Or should we just bracket the law with a *ceteris paribus* clause? Cartwright exhorts us to put this problem into different terms. Instead of speaking in terms of laws and exceptions to laws, think in terms of capacities. The regularity exhibited in the first statistical analysis points to a capacity that exists even when the background conditions have shifted. Only now the capacity no longer reveals itself in the probabilities because there are other capacities to distort its presence. Perhaps it is best to use Cartwright's own words sum to up her argument:

Nature, as it usually occurs, is a changing mix of different causes, coming and going; a stable pattern of association can emerge only when the mix is pinned down over some period or in some place. Indeed, where is it that we really do see associations that have the kind of permanence that could entitle them to be called law-like? The ancient examples are in the heavens, where the perturbing causes are rare or small in their influence; and the modern examples are in the physics laboratory, where...our control is so precise that we ourselves can regulate the mix of causes at work. Otherwise, it seems to me, these vaunted laws of association are still very-long-outstanding promissory notes: laws of association are in fact quite uncommon in nature, and should not be seen as fundamental to how it operates. They are only fundamental to us, for they are one of the principal tools that we can use to learn about nature's capacities. (Cartwright 1989, 182)

Although she makes no reference to the work of philosopher of science, Roy Bhaskar, his 1978, *A Realist Theory of Science* bears a striking resemblance to Cartwright's position.⁵¹ Before getting to the heart of Bhaskar's argument, it may be helpful to introduce some of the terms and phrases that he employs. Bhasker uses the term "tendency" in much the same way as Cartwright uses "capacity." But Bhaskar also uses the terms internal structure, mechanism, and causal power. Essentially, an entity, or a 'thing', is composed of an internal structure that gives it certain causal powers. Diamonds, for instance, have a certain chemical composition that make up their internal structure, and gives them the causal power or tendency to cut glass. When the causal power or tendency is triggered, it acts as a mechanism, thereby causing the effect. Diamonds possess the causal power to cut glass at all times, but will not until triggered (e.g., by someone picking up a diamond and scraping it across the surface of a window pane).

With this background, we can now delve deeper into Bhaskar's philosophy. Bhaskar draws a distinction between causal laws and 'mere' regularities. Mere regularities operate only because humans intervene to artificially create what Bhaskar calls a *closed system*. A closed system exists when background conditions are invariant. Outside of astronomy, closed systems only occur when manufactured in the scientist's laboratory through experimental control. (Note the similarities here to the last passage quoted from Cartwright.) In *open systems*, by contrast, there is a constant flux and changing mix of

⁵¹ Bhaskar argues for what he calls *critical realism* as opposed to scientific realism.

background conditions, making regularities a rarity. But the lack of regularities does not necessarily mean that causal laws are not at work. It only means that there are *multiple* causal laws at work, making it appear as if there are no regularities. Thus, Bhaskar (33) claims: “causal laws endure and continue to operate in their normal way under conditions, which may be characterized as ‘*open*’, where no constant conjunction is forthcoming.”

Given that regularities are only observed by way of experimental control and human intervention, Bhaskar contends that Humean empiricists are at a loss to adequately explain how it is that laws identified through experimental activity are successfully applied outside of the experimental situation, where no constant conjunctions prevail. In Bhaskar’s view, critical realism can account for this seemingly peculiar situation. The answer lies in the notion of tendencies. Even when constant conjunctions do not prevail, causal laws do. That is, it is not that a law identified experimentally stops being a law outside of the experimental situation. Rather, the law continues to apply but is not empirically identifiable because other laws are operating simultaneously that confound its effects. This is why Bhaskar argues that a constant conjunction is neither a *necessary* nor *sufficient* condition for a scientific law. Bhaskar notes that it has long been argued that regularities are an insufficient condition for laws because a constant conjunction may always be spurious. But the alleged value added of his argument is to show that constant conjunctions are also not necessary: an absence of constant conjunction does not imply the absence of a causal law at work.

So far there is a great deal of symmetry between the arguments of Cartwright and Bhaskar, but Bhaskar also develops the notion of internal structure, discussed above, which gives rise to causal powers. Bhaskar's claim is that the "weakness of the Humean concept of laws is that it ties laws to closed systems, viz. systems where a constant conjunction of events occurs." Instead of tying laws to regularities, Bhaskar (14) argues that we should tie laws to tendencies that are generated from the internal structure that 'things' possess, which in turn give rise to 'generative mechanisms' once triggered:

The real basis of causal laws are provided by the generative mechanisms of nature. Such generative mechanisms are, it is argued, nothing other than the ways of acting of things. And causal laws must be analysed as their tendencies. Tendencies may be regarded as powers or liabilities of a thing which may be exercised without being manifest in any particular outcome.

According to Bhaskar it is the job of the scientist to discover the internal structure or generative mechanisms that are possessed by entities. This is in opposition to the Humean empiricist goal of identifying constant conjunctions between events: For Bhaskar, "The world consists of things, not events." Most 'things' are complex objects, in virtue of which they possess an ensemble of tendencies. And it is by reference to the exercise of their tendencies that the phenomena of the world are to be explained (51). Thus, laws are not statements about events, but rather statements about "the ways of acting of independently existing....active things"(178, 52).

In their 1975, *Causal Powers*, philosophers of science Rom Harré and E.H. Madden construct a theory of science that anticipates both Cartwright's and Bhaskar's

endeavors.⁵² Like Cartwright and Bhaskar, their argument is directed explicitly against Humeans.

The central concepts behind Harré and Madden's approach are by now familiar. Entities possess an underlying structure that gives rise to certain causal powers, tendencies, or mechanisms. The most important aspect of causation is not events and regularities, but rather the "powerful particulars" (entities that possess causal powers) that give rise to these regularities. But to admit as much is to concede the notion of "natural necessity" and thus to be in direct opposition to Hume and his followers. Harré and Madden (84) note that the Humean's rejection of causal powers and natural necessity is "...due to a mistaken metaphysics in which 'power' is seen as a concept surviving from magic, an occult quality appealing only to those of too tender of mind to face the stern truth of empiricism." In fact, they counter, there is nothing occult about these concepts: "The ineliminable but non-mysterious powers and abilities of particular things...are the ontological 'ties that bind' causes and effects together...." (11).

An example offered by Harré and Madden (11) will illustrate their argument. Consider a simple water pump. To raise water up the pump and out of the spigot requires

⁵² Harré and Madden's work is the oldest attempt, of which I am aware, to develop a *comprehensive* causal power approach. But their work is preceded by some notable attempts. See, for example, Mellor (1974, 157-181) and Armstrong (1968). Of course, in pre-modern times, several philosophers adopted a causal powers type approach: these were the very philosophers that Hume was responding to in his analysis of causation.

atmospheric pressure acting on the pump's reservoir. Here we would say that the atmosphere has the ability or causal power to push water up the cylinder. This is a power that the atmosphere always possesses, and we know that the atmosphere has this power because of its internal structure. Specifically, "the atmosphere is a blanket of air around the surface of the earth. Air has weight and so exerts pressure, and the farther down in the blanket of air the greater the weight of the air above, and so the greater the pressure, etc." But this causal power possessed by the atmosphere will not always manifest itself in an empirical regularity. In the case of the water pump, there are other causal mechanisms that must be activated, such as a partial vacuum in the cylinder of the pump. This does not, however, mean that atmospheric pressure and the raising of water from a pump's reservoir are not causally related. The causal power of the atmosphere is always present but is sometimes overridden or distorted by the presence (or absence) of other causal powers.

ii. Causal Power in the Social Sciences

A number of attempts have been made to apply and extend Roy Bhaskar's critical realism to the realm of the social sciences.⁵³ In *Economics and Reality*, Tony Lawson (1997), contends that Bhaskar's critical realism helps to explain the "failure" of contemporary mainstream economics, with its reliance on the "deductivist mode of explanation." In *Realism and Social Science*, sociologist Andrew Sayer (2000, 14), contends that "causation is *not* understood on the model of a regular succession of events, and hence explanation need not depend on finding them, or searching for putative social laws."

⁵³ Bhaskar (1979) himself discusses critical realism as it applies to the social sciences.

Instead, Sayer argues, causation should be understood in terms of structures and the causal powers they yield. In “Regression Analysis and the Philosophy of Social Sciences – a Critical Realist View,” Amit Ron (2002) also draws on Bhaskar to help make sense of common methodological practices in her field. Ron’s central thesis is that when practitioners “play” with statistical data “behind closed doors,” they do not search out empirical laws, but rather try to “bring forth evidence of an otherwise hidden power or tendency of a ‘thing’.” Finally, in *Realistic Evaluation*, Ray Pawson and Nick Tilley (1997) contend that critical realism provides a philosophical foundation for significantly altering the manner by which mainstream evaluation research uses experimental methods.

Each of these aforementioned works, extend Bhaskar’s critical realism in interesting ways, and I will have occasion to detail some of the specific arguments in later chapters. For my immediate purposes, however, it is sufficient to note that critical realism – with its emphasis on causal powers – has provided the philosophical foundation for a small body of work in the social sciences. Other philosophers of social science have adopted a causal powers approach without necessarily drawing so explicitly and heavily on Bhaskar, but still relying on the pioneering work of philosophers of science such as Harré and Madden, and Cartwright.

We have seen that both Jon Elster and Daniel Little endorse a causal mechanism approach, but we will now see that both are also proponents of a causal power approach. In Elster’s 1989 *Nuts and Bolts* we see the seeds of a substantial shift in his approach to causal mechanisms, which matures in his 1998, “A Plea for Mechanisms.” In this later

approach, causal mechanisms are no longer fine-grained processes that are housed in the 'black box' between cause and effect, but rather represent probabilistic relationships.

Elster claims that laws have the form, "If conditions C_1, C_2, \dots, C_n , obtain, then *always* E," whereas mechanisms have the form, "If C_1, C_2, \dots, C_n , obtain, then *sometimes* E." Elster (1998, 47-48) explicitly notes the change in his approach to causal mechanism:

In *Explaining Technical Change*, I used the term "mechanism" in a sense that differs from the one I adopt here...In that earlier analysis, the antonym of a mechanism is a black box...In the present analysis, the antonym of a mechanism is a scientific law.

Although Elster does not use the language of causal power and internal structure, his approach is very much related. His new definition of mechanism is "the antonym of a scientific law," because he is now referring to the capacity or power of a 'thing' that is *always* present but only *sometimes* exercised.

Daniel Little's 1998 *Microfoundations, Method, and Causation* explicitly argues that social scientists should adopt scientific realism with its emphasis on causal powers. As we saw earlier, Little's 1991 *Varieties of Social Explanation* defended a causal mechanism approach and in this later work he continues to use the term causal mechanism. Only this time Little draws on Cartwright; and his use of causal mechanism does not refer to a process between events but rather to the activation of a causal power that is possessed by a social entity.

Key to understanding Little's approach is his distinction between *governing* and *phenomenal* regularities. Governing regularities, as the name would suggest, govern – i.e., produce a certain kind of result from above. The notion of a law of nature represents a paradigm of a governing regularity: a description of the laws that generate the behaviour of a given kind of 'thing.' For example, it is a governing regularity that the forces described by electrodynamics attract protons and electrons. A phenomenal regularity, by contrast, is a regularity of behaviour that emerges from the real causal properties of a thing, but which does not itself give rise to, or constrain, the thing's behaviour. For example, it is a phenomenal regularity that glass flows slowly; given the real constitution of glass, it emerges that glass has many of the phenomenal properties of a liquid (240).

Little contends that in the social sciences there are no governing regularities to be found. There are, however, phenomenal regularities; and as suggested by the example, these phenomenal regularities are the consequence of the internal structure of social entities. Further, the internal structure of these social entities can always be reduced down to the level of individuals. If there are governing regularities, these are the regularities of individual agency, such as the principles of rational choice theory or the findings of motivational psychology. But social regularities emerge; they do not govern. Hence, Little's emphasis on "microfoundations."

An example that Little draws from the social sciences illustrates his argument. Various laws of the modern state have been developed, including that states maximize revenues

and state crises cause revolutions. But these regularities are phenomenal regularities, not governing regularities, because they are the “product of a number of agents whose purposes, powers, and opportunities are similar in many different social contexts...” That is, there are “underlying institutional and individual-level circumstances that give rise to the regularities of state behaviour” (241).

For Little, this distinction between governing and phenomenal regularities makes it clear that social scientists should *not* rest content with the discovery of phenomenal regularities. This is because in and of themselves, phenomenal regularities are not explanatory. “If we want to know why windows are thicker at the bottom, it is not explanatory to offer the argument that windows are made of glass and glass flows like a liquid. Rather, we want to know what it is about the fine structure of window glass in virtue of which it flows.” Likewise, the social ‘law’ that state crises cause revolution is not in itself explanatory. To make it explanatory we need the causal power that gives rise to the phenomenal regularity, which in turn requires that we know the internal structure and causal powers of the social entity in question. Moreover, once we have identified the causal power, we no longer require the phenomenal regularities to explain (242-243).⁵⁴

iii. Conclusions: Pragmatic Implications

⁵⁴ I would argue that Little’s “glass” example and his “revolution” examples are not comparable. I agree with Little that a correlation between state crises and revolution is not explanatory. But to explain the thickness of glass in terms of how glass flows is, in fact, explanatory.

Unlike proponents of the causal mechanism approach, causal power proponents offer little in the way of pragmatic advice as to how, exactly, unobservable causal powers can be determined. On what basis can causal power be inferred? What analytical techniques or methodologies can be employed to determine the internal structure of a ‘thing’?

Regardless, for my purposes these types of questions are moot. I begin the construction of the SSR approach to causation in Part II by arguing that the concept of causal power has serious philosophical flaws. I contend that causal mechanism is a much more useful and relevant concept. Consequently, I adopt the concept of causal mechanism rather than causal power as the basis for the SSR approach. It is to these issues that I now turn in Part II.

Part II: The Social Scientific Realist Approach

In Part II of the thesis, I attempt to develop the philosophical and methodological bases for my own approach to causation: the SSR approach. In Chapter V, I expose philosophical flaws in the causal power approach and argue for the superiority of an approach based in causal mechanism. In Chapters VI and VII, I define “mechanism.” I also discuss three distinct ontological properties of social mechanisms (physicalness, agency, and intentionality) and describe how three corresponding methods (process-tracing, rational choice modeling, and interpretation) generate findings based on these properties that explain regularities in unique ways. Chapter VIII shows that even if the social world is comprised primarily of mind-dependent *ideas*, mechanisms in the social world can be said to have a truth value. In Chapter IX, I argue that mechanisms can always be explained in terms of deeper processes, the implications of which are discussed in Chapter X. Chapter X, the climax of Part II, weaves together the disparate features of mechanisms discussed in Chapters VI through IX. I explain how the argument from coincidence (AfC) can help establish whether or not mechanisms function as theorized and whether or not they are literally true, and I discuss the methodological implications of the SSR approach for social scientific research. In addition, Chapter X refines the debate over causation by identifying four philosophical positions that differ based on their respective beliefs about the epistemological and ontological status of theoretical mechanisms.

Chapter V: Why Causal Mechanism Trumps Causal Power

Before I begin my argument for the integral role that mechanism plays in making causal inference, it is necessary to explain why I do *not* focus my approach on the concept of causal power. In this chapter, I argue that the causal power approach is philosophically flawed, and that the concept of causal mechanism provides a superior base from which to launch a challenge against the black box approach to causation.

Causal power proponents do not champion their approach only as a replacement to Humean causation: they also claim that causal power is more fundamental to an understanding of causation than causal mechanism. Recall that for proponents of causal power, a causal mechanism merely captures the *activation* of a causal power. Volcanoes have the power to erupt but only do so when certain geological and chemical conditions prevail. When the volcano is activated, the causal power becomes a mechanism. In this sense the causal power precedes the mechanism temporally and works at a higher level of generality. In short, it is more fundamental to an understanding of causation.⁵

⁵ It may seem odd that I would pit proponents of these two approaches against one another when previously I have noted that the literature often glosses over or else conflates the distinction between causal mechanism and causal power. I do not mean to suggest here that causal power proponents are necessarily cognizant of their claim to superiority. I only claim that cognizant or not, the concept of causal power is considered to be more profound by its proponents than is the concept of causal mechanism. It is worth noting on this point that both Daniel Little (compare Little 1991 and Little 1998) and Jon Elster (compare Elster 1983 and Elster 1998) converted from

In this chapter, I challenge both claims made by proponents of the causal power approach. I contend that the approach is subordinate to the causal mechanism approach and that causal power is an inappropriate concept upon which to build a challenge against the Humean black box approach.

The first section of this chapter demonstrates that the causal power approach cannot provide satisfactory answers to causal questions, an inability that does not impede the causal mechanism approach. The second section argues that the inability to answer causal questions is a function of the level of generality at which causal powers operate. Finally, in the conclusions, I briefly relate my arguments to Chapters VI and VII, which distinguish three ontological properties of mechanisms.

i. Answering Causal Questions

The ability to provide answers to causal questions is a key aspect of any theory of causation. Causal questions often mark the starting point for research into causal relationships in the natural and social sciences. It is natural to follow up an observation *of* a phenomenon with a question about what factor(s) might *cause* that phenomenon.

the causal mechanism approach to the causal power approach, not vice versa. And Dessler (1991), who endorses the causal mechanism approach, contends that deeper explanations for his reasoning can be found in the concept of causal power.

Unfortunately, the causal power approach faces at least three drawbacks in its attempt to provide answers to causal questions: (1) it sometimes provides incomplete and unsatisfactory answers to *singular* causal questions; (2) it sometimes fails to provide any adequate answer whatsoever to *general* causal questions; and (3) it sometimes provides ambiguous answers to causal questions, whether specific or general. I show that each of these problems is remedied by thinking in terms of causal mechanism rather than causal power.

Singular Causal Questions

My first criticism of the causal power approach is that it can fail to yield complete and satisfactory answers to singular causal questions. By a complete and satisfactory answer to a causal question I mean an answer that does not beg further questions *at the same level of analysis*. For instance, suppose I ask you why you did not go to work today, and you respond that you could not walk with a cast on your leg. I might reasonably ask how you had come to break your leg, in order to provide me with a *more* satisfactory and complete explanation for why you did not go to work today. Notice that I am not asking for an explanation at a deeper level of analysis, say by inquiring into why the doctor decided your particular type of break requires a cast.⁵⁶

Sometimes the causal power approach simply will not provide a satisfactory answer to a singular causal question because it tells only part of the story – and usually the less

⁵⁶ I specify “same level of analysis,” because as we will see in Chapter IX, all explanations are in want of ever-deep explanations at lower levels of analysis.

interesting part. Consider first this non-singular question: “What causes humans to die?” (I continue the philosophers’ tradition of using macabre examples.) The causal power proponent instructs us to look at the internal structure of humans and understand how the vital organs of the human body – such as the lungs, brain and heart – can all induce death if they cease to function.

But what if the causal question is singular? “What caused *John* to die?” One can imagine the parents of John arriving at the hospital to be briefed by a policeman and a doctor as to the cause of their son’s death. The doctor speaks first, explaining to the family – in a manner that a causal power proponent would advocate – the intricacies of human anatomy, hauling out textbooks to demonstrate the pulmonary system, and using a chalkboard to diagram the flow of blood. This type of explanation makes perfect sense according to the causal power approach, given that John is a human and human death must be explained by the internal structure of humans.

But is this type of explanation really satisfactory to the bereaved parents? I suspect that the policeman’s explanation of John’s death – one that a causal mechanism proponent would advocate – might prove more enlightening, if also more horrific: “John was driving in the right hand lane at approximately 110 Km/h without his seat belt properly fastened. He ran over a sharp piece of scrap metal dropped from a truck that had passed by moments earlier. The left front tire of John’s car blew out causing the vehicle to swerve into the highway’s median. John was thrown from the vehicle into oncoming traffic on the opposite side of the highway, whereby he was run over by a cement truck

and three Sport Utility Vehicles (SUVs). In the process of being thrown from the vehicle and being run over, he lost the function of all vital organs.”

John’s death is explained to his parents in terms of a causal process, beginning with the presence of a piece of scrap metal on the road and ending with the failure of John’s vital systems. It is not that the doctor’s explanation regarding the failure of John’s vital systems is wrong. It is simply that it represents only one aspect of a complex process and makes for a very unsatisfactory explanation.⁵⁷ With the causal mechanism approach, conversely, we have a complete explanation that appalls John’s parents, but at least provides them with closure. Thus we can solve this problem of incompleteness by thinking in terms of causal mechanism rather than causal power.⁵⁸

It may be objected by the causal power proponent that in order to *explain any singular* event we are required to list the ‘background conditions’ (additional causal powers) that allow for the *activation of the causal power in question*. So, for example, the causal

⁵⁷ It may be objected that a causal power proponent would point in the example at hand to the causal power of scrap metal, cement trucks, and/or, SUVs. Unfortunately, this solution runs headlong into another problem with the causal power approach discussed below: namely, ambiguity as to what ‘thing’ gets imbued with causal power in answering a causal question. Moreover, as I point out below, the causal power approach is lacking because it does not generate a sequential explanation.

⁵⁸ The problem that I raise could be thought of merely as an issue of differing interests or subjectivity. See footnote 61, below.

power proponent explains the explosion of a particular piece of dynamite by noting that the causal power of combustibility was activated by the lighting of the wick and the presence of dry air. In the case of John's death, the causal power of his vital organs to shut down was activated by various 'background conditions', including his being run over by three SUVs. However, the causal power approach does not necessarily provide any *sequence* to these conditions. The causal mechanism approach, by contrast, provides an orderly sequence of events by following *through time* the events that lead from cause to effect.⁵⁹

General Causal Questions

I say that the causal power approach might provide only an incomplete and unsatisfactory causal account, but this might be an overly generous criticism. In fact, sometimes the causal power approach provides no causal account whatsoever, especially when used to answer general causal questions. Consider the question, "What causes water to boil?" Following the prescription of the causal power approach, we are told to answer this question by looking to the internal structure of water. Presumably, we would find that the chemical composition of water is such that, when the conditions were right, water would boil. But this advice proves to be very misleading in the case at hand.

⁵⁹ As noted above, an activated causal power is referred to as a 'mechanism'. But we can see from this example that the causal mechanism approach uses the term, "mechanism," to refer to the *entire* process – not merely one aspect of the process, as does the causal power approach.

Take a simplified version of why water boils. When water is heated, the speed of the molecules in the liquid quickens, causing an increase in vapour pressure. When vapour pressure becomes equal to atmospheric pressure, molecules are regularly ejected into the air above the liquid. These ejected molecules take the form of bubbles. Atmospheric pressure then, is an integral component to an understanding of why water boils. The internal structure of water *is* important, but only *relative* to atmospheric pressure. In and of itself, the internal structure of water *does not* explain why water boils. It would *not* suffice to say that what causes water to boil is an increase in vapour pressure.⁶⁰

An answer to a causal question that invokes only the internal structure (causal power) of water would be more than incomplete – it would be misleading. It would be misleading because two pots of water might have an identical increase in vapour pressure, but due to differing atmospheric pressures only one of these two pots might boil. At the top of Mount Everest, it will take a much smaller amount of vapour pressure to boil water than it will at my home in Toronto. But the causal power approach fails to handle this case. Causal power proponents cannot retort that atmospheric pressure is merely a background condition – something that has to be ‘right’ in order for the internal structure of water to unleash its causal power. There simply is no objectively ‘right’ atmospheric pressure – there is no atmospheric pressure that will allow water to boil every time. Instead, the question of whether water will boil or not, depends on the atmospheric pressure *relative*

⁶⁰ It would also render the causal power approach vapid if one were to invoke the causal powers of both water and atmosphere: It simply runs counter to the parsimony and universality that the causal power approach is supposed to contain.

to the vapour pressure of the water. Both factors, vapour pressure and atmospheric pressure, are integral to a response to our general causal question, “What causes water to boil?”

The causal mechanism approach is not constrained in the same way because it intrinsically allows for “relativized” explanations – e.g., by following the process from the heating of the molecules through to their interaction with the atmosphere above. A problem only arises when attempting to “locate” the “causal action” in one particular ‘thing.’ In the example at hand, the causal power approach puts all the “causal action” in the water itself, thereby failing to provide an adequate causal explanation.

Ambiguity as to what gets the causal power

A third problem with the causal power approach is that, in answering singular or general causal questions, it can be entirely ambiguous as to what ‘thing’ in a causal question is imbued with the corresponding causal power. Take a slight alteration to an example provided by Cartwright (1989). Cartwright poses the question, “Why does Aspirin causes the relief of headaches in the human body?” Her response: Aspirin, by virtue of its internal structure, has the capacity to do so.

But why not attribute a causal power to the human body to relieve itself of headaches? It is perfectly consistent with the causal power approach to maintain that the human body, by virtue of its internal structure, always carries the causal power to relieve headaches. The headaches will not always be relieved, but given the right background conditions

they will be. One such background condition would be the ingestion of Aspirin. Are we then to assume that *both* Aspirins and the human body have the causal power to relieve headaches? Citing two ‘things’ with the same causal power hardly seems like a satisfactory answer to a causal question; it is even worse to choose one ‘thing’ over another without warrant.⁶¹

By extension, I would argue that almost anything could be imbued with almost any causal power. An axe may be said to have the causal power to kill people. All that is required is the appropriate background conditions – e.g., that the axe be wielded in an inappropriate and dangerous fashion. This line of reasoning quickly leads to the truly absurd – filing cabinets have the causal power to stop highway traffic (as surely they do if two or three fell off the back of a truck) and credit cards have the causal power to jimmy

⁶¹ The objection I raise here can be seen as a problem of interests, or subjectivity: it is up to the researcher to determine what ‘thing’ he or she will choose as an answer to a causal question. This situation resembles a well-known example of the subjectivity of causation provided by Collingwood (1940, 304): “...A car skids while cornering at a certain point, strikes the [curb], and turns turtle. From the car-driver’s point of view the cause of the accident was cornering too fast, and the lesson is that one must drive more carefully. From the county surveyor’s point of view the cause was a defect in the surface or camber of the road, and the lesson is that great care must be taken to make roads skid-proof. From the motor-manufacturer’s point of view the cause was a defective design in the car, and the lesson is that one must place the center of gravity lower.” Collingwood’s example points out that different researchers will choose different comparison situations, depending on their research interests.

doors (as they do if the door locks are particularly loose).⁶² These examples may be absurd, but the absurdity is only licensed by the argument for causal powers.⁶³

Again, to rectify this shortcoming, it is necessary to think in terms of causal mechanism rather than causal power. What is really important and relevant about a causal power is the contextual conditions that activate this power. How does a filing cabinet cause traffic to stop? The answer to this causal question will surely not be provided by an internal description of filing cabinets, as large objects constructed of metal. Or rather, this description will provide only part of the answer to our question: again, the less interesting part. It is true that no driver would want to drive headlong into a filing cabinet, so an

⁶² It may be objected that it is not filing cabinets or Aspirin per se that have causal powers but rather the properties attributed to them. Thus, it is, respectively, the property of being large and heavy, and the property of acetylsalicylic acid (ASA) that contain the causal powers to stop traffic and relieve headaches. Although causal explanations based on properties reduce the absurdity of the causal power approach it cannot eliminate it entirely. This is because it *remains* ambiguous as to what ‘thing’ is ultimately responsible for a causal event or phenomena: too much depends on the activation of background conditions. Thus, ASA has the causal power to relieve headaches, but so too does the human body provided that it ingests a product containing ASA.

⁶³ Comedian, Steve Martin (1998, 108), satirized this absurdity: “Universal Causality: This is the law that has the legal world most excited. It rests on the proposition that “anything can cause anything,” or, more simply put, the “Bill Gates gave my dog asthma” principle. If the law of Universal Causality bears out, the economy will receive an invigorating boost when everyone sues everyone else for everything. Everything actionable that ever happened to you will be the fault of your neighbour, who, in turn, will sue Bill Gates, who, in turn, will sue himself.”

understanding of its composition is important. But more interesting are the background conditions that led to the activation of the causal power. For instance, we can imagine a filing cabinet that lies dormant for years, gathering dust in a company's warehouse. At some point the filing cabinet is requested by head office, and a truck is commandeered to move the filing cabinet to its new destination. On the way, the filing cabinet shakes loose from the truck, falling onto the road, causing cars to slam on their brakes. Consequently, traffic grinds to a halt. This description of why a filing cabinet can cause traffic to stop describes a process; and this process is best captured by the concept of causal mechanism.

ii. The Quest for Genuine Universality

The three problems I have identified so far – incompleteness, lack of explanation, and ambiguity – point to a general difficulty with the causal power approach. Recall that for causal power proponents, causal powers operate at a higher level of generality than do mere causal laws bracketed by *ceteris paribus* clauses. Cartwright, for instance, argues that a *genuine* law should be characterized by “contextual unanimity” – that is, a cause should raise the probability of its effect regardless of prevailing background conditions. But I contend that the price exacted for this level of generality is high, indeed too high to make the concept of causal power of any value whatsoever. It turns out that what really makes for an interesting causal explanation is not the mere causal power – which exists at all times – but rather the mechanism that exists only when the causal power is activated (i.e., when the causal power is set in motion by the appropriate conditions).

Indeed, I question the entire effort to generate truly universal causal laws. This is important because it is by virtue of this line of reasoning that the ground is laid for the causal power approach. That is, causal power proponents criticize Humeans for establishing a requirement for laws – that they be genuinely universal – and then failing to provide a theory of causation that meets this requirement.⁶⁴

But clearly it is false to suggest that Humeans and other positivists have ever tied laws to “exceptionless” regularities. Recall from Chapter II that in his famous billiard ball example, Hume himself states: “Let us try any other ball of the same kind in a like situation...” The key phrase here is, “*in a like situation.*” In other words, Hume recognized that constant conjunction would not be constant over all background conditions and that this was not necessary to establish a law. Hempel (1959) as well explicitly and clearly recognizes the parameters of laws:

[I]f we are to predict, by means of the laws of classical mechanics, the state in which a given mechanical system will be at a specified time t , it does not suffice to know the state of the system at some earlier time t_0 , say the present; we also need information about the boundary conditions during the time interval from t_0 to t , i.e., about the external influences affecting the system during that time... This shows that even the laws and theories of the physical sciences do not actually enable us to predict certain aspects of the future

⁶⁴ Keat and Urry (1975, 14), for instance, state that for Humeans, scientific laws “must not be restricted in their application to any finite region of space and time: they must hold true for all times and places.” See also, Cartwright (1989, 143-45).

exclusively on the basis of certain aspects of the present: the prediction also requires certain assumptions about the future. But, in many cases of nomological prediction, there are good inductive grounds, available at t_0 , for the assumption that during the time interval in question, the system under study will be practically “closed”, i.e., not subject to significant outside interference...⁶⁵

It appears as if causal power proponents have erected a straw man argument: Humeans do not tie causal laws to universal generalization or contextual unanimity. Recognizing this fact would seem to undercut the causal power enterprise, which is promoted as an attempt to rectify the principle of genuine universality that Humeans safeguarded so poorly.

iii. Conclusions: A Deeper Look at Mechanisms

I have argued in this chapter that the causal power approach suffers from serious philosophical flaws, and therefore constitutes an inappropriate position on which to base a SR approach to causation. I have also suggested in this chapter that the causal mechanism approach is not hampered by these same philosophical problems.

But the causal mechanism approach, as advocated by SRs in the social sciences, suffers from some potential problems. Causal mechanism proponents in the social sciences

⁶⁵ This quotation is instructive for its use of the term “closed”, which seems to establish the ‘open-closed’ distinction that critical realists claim to originally introduce and that plays such a key role in their lexicon. The behaviorist psychologist, Clark Hull (1966 [1943], 2), discussed in Chapter XI, also explicitly recognizes the open/closed distinction.

sometimes fail to clearly delineate the properties of mechanisms that produce very different types of explanatory mechanisms.⁶⁶ Beyond case study methods, they also sometimes fail to adequately address how mechanisms are tested and fail to recognize the central importance of the AfC to SR. Further, causal mechanism proponents sometimes fail to adequately address the ontological dimension of mechanisms. The remainder of Part II is therefore devoted to remedying these and other shortcomings in existing causal mechanism approaches.⁶⁷

⁶⁶ The examples that I use in this chapter relate for the most part to a specific property of social mechanism: namely, physicalness, which is characterized by chains of events and phenomena. This property of mechanisms is used here because it is highly intuitive and applies equally well to the natural and social scientific examples that I draw upon.

⁶⁷ Not all existing SR-based causal mechanism approaches, of course, suffer from all of these problems. See Chapter XIV for more details on the “value added” of the SSR approach developed in this thesis.

Chapter VI: Physicalness and Agency

Now that I have argued against causal powers and for causal mechanisms, it is time to get more specific about the composition of social mechanisms. In this chapter and the next, I contend that mechanisms of social causation consist of three distinct ontological properties: physicalness, agency, and intentionality. Each of these three distinct ontological properties can be identified by three corresponding social scientific methods: namely, process-tracing, rational choice modeling, and interpretation. Each of these methods, in turn, can generate findings that explain regularities in a unique fashion. These relationships are provided in Table 1.

The first ontological property of social mechanisms is physicalness. Physicalness here refers, in part, to the material or corporeal, but physicalness *is not* limited to the observable realm. As will be discussed in Chapter VIII, SR holds that unobservable entities and ideas can be just as *real* as observable entities and artefacts. What is most important about physicalness from my perspective is the notion that ‘things’ in this world – whether material or ideational – impact upon one another in a mechanistic fashion, much as billiard balls impact upon one another on a billiard ball table, or molecules inside a gas impact upon one another.

Process-tracing identifies this property of social mechanisms and generates findings that explain regularities as series of linked events or phenomena that exist and operate between a cause and its effect: that is, an intervening chain. This physical explanation of

regularities is highly intuitive and common to both the natural and social sciences. Therefore an explanation of physicalness and its associated methodology of process-tracing occupies only a small portion of the chapter.⁶⁸

What is contentious about physicalness, is whether it can properly be described as an ontological property of *mechanisms*. Humeans sometimes contend that an intervening chain of events or phenomena is nothing more than a series of regularities. Thus, if the causal relationship between ‘A’ and ‘D’ is explained by a series of regularities linking ‘A’ to ‘B’, ‘B’ to ‘C’, and ‘C’ to ‘D’, mechanism is a superfluous concept: the concept of regularity is all that is required.⁶⁹ My response to this counterargument against mechanism will be supplied in Chapter X, where I explain the role that mechanism plays in causal analysis. For now, it is worth noting that if the second and third properties of mechanisms that I identify below (agency and intentionality) have merit, then social mechanisms do not consist merely of regularities, and therefore irrespective of physicalness, the concept of mechanism is not superfluous.

The second ontological property of social mechanisms is agency. Agency refers to preferences and actions on the part of humans or collectivities of humans (such as a

⁶⁸ In fact, some psychologists (e.g., Ahn et al. 1995, Koslowski 1996) contend that humans intuitively make causal judgements in terms of intervening chains.

⁶⁹ We will see in Chapter X that this argument holds for both HEs and HRs.

corporations, clubs, or states). Rational choice modeling⁷⁰ identifies this property and generates findings that explain regularities as the aggregation or strategic interaction of individuals' preferences. Although political scientists often use the findings generated by rational choice modeling to explain regularities, this technique is not necessarily well understood or overly intuitive. Therefore, I spend more time explicating how rational decisions can yield regularities than I do explaining how a chain of events or phenomena can yield regularities.

The third and final ontological property of social mechanisms is what philosophers call intentionality. Intentionality, according to John Searle (1998, 85) are those "subjective states [that] relate [a person] to the rest of the world." These subjective states include, "beliefs and desires, intentions and perceptions, as well as loves and hates, fears and hopes." As Searle (85-6) notes, intentionality "is an unfortunate word," because it suggests a "sense of directedness" as in "I intend to go the movies tonight." However, as Searle points out, intention in this ordinary sense is "just one form of intentionality."

⁷⁰ Rational choice research is based on certain methodological techniques as well as theoretical postulates. I use the term "rational choice modeling" as opposed to the more common "rational choice theory" to emphasize this former characteristic. But I do recognize that theoretical postulates undergird rational choice's methodological techniques: if one adopts other psychological theories of agency, then rational choice modeling may be inapplicable. This issue is taken up further below.

I will focus in this thesis on a particular form of intentionality that Searle (118-21) calls “collective intentionality.” *Collective intentionality* is of the form, “we intend,” “we believe,” and so on. As Searle (118) points out, thinking in terms of “we” does not remove the “I” from the equation: “Now of course, if I have a ‘we intention,’ I must also have an ‘I intention,’ because if I am intentionally doing something as part of our doing something, then I must intend to do my part. And in order to intend to do my part, I must intend that I do something that is part of our doing something.” Thus, this third ontological dimension of social mechanisms does not entirely remove subjective beliefs and desires and individual human action from the equation; rather it takes these individual beliefs, desires and actions to be intimately tied to, and ultimately directed by, collective beliefs and desires.

Unlike agency, collective intentionality suggests that the basis of social action is not rational calculus but rather shared purposes and common meanings. Social actors are not necessarily cognizant of their own intentions: societal-level forces that cannot be reduced to conscious decisions motivate them. In opposition to rational choice modeling, interpretation holds that collective intentionality “...cannot be just postulated...but must be investigated, for [it] can be complicated and obscure and may even be denied by those to whom [it is] rightly attributed” (Forbes 2004, 65).

Intentionality, especially collective intentionality, is closely tied to the method of interpretation (Forbes 2004, 65, George and Bennett 2005, 129). Interpretation identifies this ontological property of social mechanisms and generates findings that explain

regularities as being *constituted by* collective intentionality, or what proponents of interpretation more often call *intersubjective meanings*. Intersubjective meanings produce a social context that “make possible” or “enable” the existence of regularities.⁷¹

Explaining regularities by intersubjective meanings is far less familiar to most political scientists than explaining regularities as the product of an intervening chain of events or phenomena, or as product of individual decisions. I therefore devote the entire following chapter (Chapter VII) to explaining intersubjective meanings and their relationship to regularities. Moreover, many proponents of interpretation will reject outright the notion that interpretation is in the business of causal analysis. This is unlike the situations with process-tracing and rational choice modeling, which are often both accepted as being amenable to causal analysis. I therefore spend a large portion of Chapter VII defending the compatibility of interpretation and causal analysis.

The first section of this chapter provides a general definition of “mechanism” for the social sciences. The second section discusses the relationship between physicalness, process-tracing, and regularities, while the third section discusses the relationship between agency, rational choice modeling, and regularities. In each case I provide several examples from social scientific scholarship, including recent political science scholarship. The conclusions set the stage for my discussion, in Chapter VII, on the relationship between intentionality, interpretation, and regularities.

⁷¹ By the same token, intersubjective meanings can “make *impossible*” the existence of regularities.

i. Defining Mechanism

SRs in the social sciences broadly agree on what mechanisms are: namely, they are things that can explain – account for – regularities, causal relationships, or phenomena.

Unfortunately, this broad agreement on what mechanisms are, has not translated into agreement on a precise definition of mechanism. Copious amounts of ink have been spilled in attempting to provide a definition of mechanism that captures what is a fairly intuitive concept. Table 2 displays a list of thirty such definitions, as compiled by Mahoney (2001, 579-80).

There are three strands running through the definitions listed in Table 2 that I reject.

First, mechanism is sometimes defined in terms of causal power as per critical realism.⁷²

This is seen in the definitions of Keat and Urry, Harre, and Steinmetz. As I argued in Chapter V, the concept of causal power is philosophically flawed. Consequently, I reject defining mechanism in terms of causal power.

Second, some SRs in Table 2 define mechanism as a theory. This is evident in the definitions provided by Pawson, Koslowski, and Stinchcombe. Equating mechanism with theory is problematic because “theory” is too broad a term. In mathematics, a theory refers to a body of principles. In the sciences, a theory can refer to a proposed explanation whose status is still conjectural. There are important differences between these uses of the term “theory,” and that meant to be conveyed by the term “mechanism.”

⁷² As explained in Chapter IV, critical realism defines mechanisms as activated causal powers.

I conceive of mechanisms as more specific than theories: that is, they are (often) specific instances of theories in action.⁷³ For instance, *systems theory* posits that social phenomena can be explained by the relative positions that social units or entities occupy vis-à-vis one another (Parker et al. 2003, 139-43). Neorealists (in international relations) identify a particular instance of this theory in action: the relative capabilities that individual states possess vis-à-vis one another constitute a *mechanism* that constrains their behaviour (Waltz 1979).

The third and final problematic strand running through the definitions provided in Table 2 is that mechanism is overly restrictive. In particular, a number of authors – including Cowen, Gambetta, Hedstrom and Swedberg, and Schelling – define mechanism strictly in terms of rational choice modeling or some other individual-level decision-making process. This methodological individualist approach is problematic to the SSR approach, because it holds that social mechanisms have three ontological properties, only one of which is associated with rational choice modeling.

Once we remove those definitions from Table 2 that equate mechanism with causal power, with theory, or with individual-level decision-making processes, what remains is the intuitively appealing concept that mechanisms are simply explanations of regularities. The definitions provided by Goldthorpe and Elster, for example, have this intuitive appeal.

⁷³ I do not argue that this approach applies to all theories. Some theories, such as the *theory of relative deprivation*, may be sufficiently specific to be considered a mechanism.

Although tempting to define mechanism in the manner of Goldthorpe and Elster, I make my own definition of mechanism somewhat more precise by tacking on the manner in which mechanisms function. My definition of mechanism, therefore, is as follows:

Something that explains causal relations by appealing to: intervening chains of events or phenomena, the aggregation or strategic interaction of individual-level decisions, or constitutive intersubjective meanings.

ii. Physicalness, Process-Tracing, and Regularities

Social mechanisms are comprised, in part, of a chain of events or phenomena that connect cause and effect. This is sometimes referred to as a mechanistic or billiard ball model of explanation. Process-tracing⁷⁴, a method whereby in-depth knowledge of the social, economic, and political makeup of individual cases⁷⁵ is used to trace a sequence of events through space and time, often identifies this property of mechanisms in the social

⁷⁴ Some case study methods – such as cross-case and within-case comparative methods – straddle the divide between process-tracing and statistical analysis. I do not explicitly discuss comparative methods in this thesis, although I acknowledge that they can play an important role in causal analysis. For an in-depth treatment of comparative methods, see George and Bennett (2005) and Ragin (1987). For an in-depth treatment of process-tracing in particular, see George and Bennett (2005, 205-32).

⁷⁵ I adopt George and Bennett's (2005, 17-8) definition of a "case" as an instance of a class of events, such as revolutions, types of governmental regimes, kinds of economic systems, etc. See Gerring (2004) for a useful discussion on defining cases.

sciences.⁷⁶ Historians engage in process-tracing when for example, they describe the sequence of events leading up to the fall of Rome or the outbreak of WWI. Political scientists routinely use process-tracing to explain regularities and modify existing theories (George and Bennett 2005).

For instance, Sambanis (2004) draws on 21 case studies, to explain certain regularities relating to the onset of civil war, such as the correlation between state strength and civil war.⁷⁷ Sambanis argues that, in contrast with strong states, when weak states act in a decisive and repressive manner against potential internal threats, more opposition and violence is likely to follow. States will typically attempt to thwart internal threats through accommodation, but when that fails, repression usually follows. State repression, in turn, can generate intolerable social and economic conditions on the societal group that represents a threat to the government, engendering new rounds of more severe violence. For instance: “In Burundi [in the 1960s], the government excluded Hutus from elite positions and inflicted violence on their leaders. Over time, this repression led to fewer educational opportunities and less economic power for Hutus – but it eventually backfired, resulting in a large-scale Hutu rebellion that entailed Hutu coup attempts, Tutsi countercoups, [and] Hutu massacres of Tutsi in 1965 and 1972...”

⁷⁶ George and Mackeown (1985) coined the term, process-tracing, although Diesing (1971), Dray (1957), Hempel (1965), McClelland (1975), and Tilly (1997) – to name just a few – develop similar ideas.

⁷⁷ Sambanis uses the *structured-focused comparison* method (see George 1979) and process-tracing. See Sambanis’ discussion on research design, 262-3.

Thus Sambanis (271) concludes, “Case studies can give us a better sense of the dynamics of conflict escalation by presenting a sequence of events – a series of actions and reactions – linking several independent variables together in a process that culminates in war.”

Runkle (2003) provides another example of the use of process-tracing to explain a regularity. Runkle (iii) notes that “[a]lthough...quantitative studies...over the past 25 years show a strong, positive correlation between arms races and the outbreak of war...these studies tell us little about precisely that which policymakers seek to understand – the causal mechanisms that lead from arms races to armed conflict.” To that end, Runkle conducts process-tracing on four historical cases – France v. Germany, 1872-1893; Germany v. France/Russia, 1910-1914; United States v. Japan, 1916-1922; and United States v. Japan, 1934-1941 – with the aim of testing and elaborating a selection of competing causal mechanisms. More specifically, Runkle uses process-tracing to test whether arms races are more likely to engender “conflict spirals,” “economic burdens,” “false optimism and bellicosity,” “windows of opportunity,” “militarism,” or some combination thereof. “Case studies,” contends Runkle (51) can better determine whether a relationship is causal or spurious by identifying which of several possible causal mechanisms are at work.”⁷⁸ Runkle concludes from his analysis that arms races most often lead to war by creating a “windows of opportunity” that lead states to launch preemptive strikes: “[A]rms races lead to war when they create a shift in

⁷⁸ Runkle combines case study research with correlative analysis to assess the causal relationship between arms races and war.

the relative military balance of power between states that causes a militaristic state to perceive a window of opportunity through which it can launch a preventive war” (285).

Homer-Dixon (1999) also uses process-tracing to elaborate the potential linkages between environmental scarcity and violent conflict. Based on a series of process-tracing case studies, Homer-Dixon determines that scarcities of renewable resources, such as water, forests, cropland, and fish, can trip a series of intermediary physical and social events and phenomena, such as decreasing economic productivity, widening social cleavages, and declining institutional capacities. For instance, his findings indicate that deforestation can generate fuelwood shortages, which in turn constrains economic productivity. Deforestation can also generate soil erosion and increased runoff rates, both of which can lead to the disruption of hydropower production and damage critical infrastructure, and, ultimately, constrain economic productivity. These intermediary events and phenomena, in turn, can interact with one another and ultimately engender social friction and violent conflict. Thus, combinations of scarcity-induced losses in economic productivity might exacerbate existing social cleavages between rival ethnic groups within a developing nation and foment violent conflict between these groups.

iii. Agency, Rational Choice Modeling, and Regularities

Social mechanisms are comprised, in part, of agency – specifically, choices made by individuals. Rational choice modeling identifies this ontological property and generates findings that explain regularities by aggregating the decision-making processes and

actions of multiple individuals in a given time and space (decision theory), or analyzing the strategic interaction of individuals' decision-making processes (game theory).

There are other agency-based psychological theories employed in political science that also have the potential to explain regularities. Making this exact point, Bennett (2003, 12) argues, "Regarding individual mechanisms, it is important that the field include not only rational choice mechanisms, but also theories on other cognitive mechanisms..." such as prospect theory, schema theory, framing heuristics, emulation, and persuasive communication. Elster (1998) points to several non-rational choice, psychological-based, mechanisms, such as "reinforcement." Nevertheless, because of its influence in the discipline of political science, I focus on rational choice modeling.

The core premises of rational choice are as follows: (1) The fundamental unit of the social world is the individual; (2) individuals pursue goals based on their preferences; (3) in pursuing their goals, individuals are able to rank order their options according to their preferences; and (4) in choosing from the array of options open to them, individuals always attempt to choose the one that best serves their preferences (MacDonald 2003, 552, Green and Shapiro 1994, 14-17, Little 1991, 39-45).

Rational choice modeling has at least two distinct branches. *Decision theory* analyzes rational choices made by individuals, choosing from amongst a set of rank ordered options that is independent of any other individuals' choices. *Game theory* analyzes rational decisions made by individuals, whereby their options are ranked and chosen in the context of rational choices made by other individuals.

The modern rational choice paradigm in political science was adopted from the field of economics: it is considered the economic approach to the study of politics. In the field of political science, the rational choice paradigm is often referred to as the *public choice* paradigm. The public choice paradigm applies the core assumptions described above to public institutions and goods, whereas the rational choice paradigm might apply to non-public institutions and goods, such as the family unit or the market (Friedman 1995, 2).

Some SRs contend that rational choice modeling can produce explanations for causal relations in the social world. Little (1991, 39), for instance, argues:

...[C]ausal explanations of social science requires some account of the mechanisms that mediate between cause and effect. The rational choice paradigm offers a general account of such mechanisms among social phenomena. If we can assume that individuals in a variety of social settings make calculating choices based on their beliefs and goals, we may be able to explain numerous social arrangements as the aggregate effects of such choices.

I provide several examples to illustrate how rational choice modeling can generate findings that explain regularities: one classic example from the field of economics, a well-known example from sociology, and four examples from political science.

A Classic Example

Garrett Hardin's, "The Tragedy of the Commons," offers a paradigmatic illustration of how rational decisions can explain a regularity. Hardin's analysis begins with a well-

documented regularity: namely, publicly owned tracts of land inexorably exhibit environmental decay. This relationship, between public ownership and environmental decay, appears causal: the two phenomena are constantly conjoined, and environmental decay follows public ownership in time. However, this apparent causal relationship lacks an explanation.

In order to explain this phenomenon, Hardin appeals to hypothetical common pasture, which is unsupervised and shared by a number of herdsman. In this commonly owned pasture, each individual herdsman concludes that the rational course of action is to add more animals to his or her own herd. The marginal benefit of adding an additional animal to the pasture outweighs the marginal cost of using the unsupervised commons. But this is the same conclusion reached by each and every (rational) individual herdsman. The resulting accumulation of animals ultimately outstrips the carrying capacity of the common pasture, and environmental decay ensues. “Each man,” writes Hardin, “is locked into a system that compels him to increase his herd without limit – in a world that is limited...Freedom in a common brings ruin to all” (Hardin 1968, 1244).

A Sociological Example

In his 1978 *Micromotives and Macrobehaviour*, Thomas Schelling set out to explain an observed regularity: As neighbourhoods in US cities become increasingly ethnically diverse, there is a tendency for these cities to also become increasingly segregated. Does ethnic diversity cause segregation? If so, how?

Little's (1991, 43) description of Schelling shows that he uses a rational choice analysis to answer these questions:

Noting the common pattern of segregation between ethnic groups in U.S. cities, Thomas Schelling attempts to construct an explanation of this in terms of a hypothesis about the preferences of individuals. "This chapter is about the kind of segregation...that can result from discriminatory individual behavior. ...It examines some of the individual incentives and individual perceptions of difference that can lead collectively to segregation" (Schelling 1978: 138). He shows that rather weak assumptions about individual preferences are sufficient to produce sharply segregated residential patterns in the aggregate. In particular, if we assume that members of each ethnic group will tolerate an ethnically mixed neighborhood up to a certain ratio and will move if the proportion rises above that ratio, in a variety of neighborhood models it emerges that the stable equilibria are those in which the two groups are sharply segregated. This aggregate result stems *not* from the fact that each person prefers to live in a segregated neighborhood but rather from the ripple effects that follow as residents in unsatisfactory neighborhoods move into new neighborhoods, thereby altering the proportions in the new neighborhood and stimulating new movement.

Political Science Examples

William Reed's "Information, Power, and War" (2003) uses a bargaining model – an offshoot of game theory – to explain an empirical regularity between power parity and war. "With few exceptions," contends Reed (633) the scholarly consensus is that pairs of states with relatively equal amounts of observable capabilities are more likely to experience conflict." "Why is there a tendency for states with an equal distribution of

observable military capabilities to engage in militarized conflict?” [T]raditional explanations for this empirical pattern,” argues Reed (633), “are unconvincing.” Reed (633) notes the importance of clarifying the causal mechanisms, as China approaches parity with the United States and Pakistan pursues nuclear parity with India.

In order to explain the regularity, Reed develops a bargaining model that links uncertainty about capabilities to militarized conflict. Rather than focussing solely on *observable* dimensions of capabilities, Reed focuses as well on *unobservable* dimensions – most notable the uncertainty that surrounds states’ estimations of their opponents’ capabilities: “In contrast to the conventional argument that the distribution of power causes conflict, I maintain that uncertainty about the distribution of power is as important a predictor of the probability of conflict as is the observed balance of power” (633).

Whereas earlier “balance of power theorists” such as Hans Morgenthau (1948) contended that uncertainty engendered by power parity tends to make states cautious and therefore chary of engaging in conflict, more recent power “transition theorists” contend that “uncertainty makes dissatisfied states more likely to use force to alter the status quo” (634). Uncertainty creates a window of opportunity for challenger states, and thus power parity is conflict enhancing rather than conflict inhibiting. Reed’s bargaining model elaborates and refines power transition theory.

More specifically, Reed demonstrates with his bargaining model that “[w]hen states are uncertain about their opponents’ capabilities, they may either overestimate or

underestimate their own bargaining leverage. Depending on how these “information asymmetries” play out, uncertainty may either increase or decrease the risk of conflict when two states engage in bargaining. If a challenger (purposely or involuntarily) overestimates its bargaining leverage and demands more than the defender is willing to concede, the risk of conflict rises. However, if a challenger underestimates the defender’s capabilities, and thus makes an inadequate opening offer that is readily accepted by the defender, the risk of conflict declines (634). States dissatisfied with the status quo are apt to misrepresent their capabilities, thereby leading to the former scenario:

States with relatively equal amounts of power, knowing the similarity between their observable capabilities and those of their opponent, have an incentive to misrepresent their unobservable capabilities in order to strike a better bargain. The paramount role that these unobservable capabilities play in the bargaining process and the clear incentive to bluff about them result in an enhanced probability of conflict. (640)

In brief, Reed’s formal model demonstrates that when rationally motivated bargaining occurs between pairs of states roughly equal in observable capabilities, uncertainty often ensues; and this uncertainty can spiral toward militarized conflict.

Robert Putnam’s (1993) much celebrated *Making Democracy Work* provides another example from political science scholarship. Putnam demonstrates that “associational density” – the number of “horizontal” linkages between people in voluntary associations such as chess clubs or neighbourhood groups – is highly correlated to a government’s

ability to govern in a responsive and effective manner. The explanation for this correlation, theorizes Putnam, is that associational density generates “social trust” or “social capital,” which helps to overcome collective action problems.

But in order to solidify this causal relationship between social capital and good governance, Putnam does not merely work toward establishing correlational evidence. Rather, he “repair[s] to the halls of Oxford” – as Tarrow (1995, 471) puts it – to develop a complex rational choice explanation for how associational density engenders effective governance. Forbes (2004, 59-60) captures Putnam’s rational choice-based explanation:

Trusting and trustworthy citizens are more able to co-operate with each other, on the basis of voluntary agreements, than are those who lack trust in each other and cannot make credible commitments. A dense (and closed) network of civic engagements sustains generalized trust because it threatens naturally self-interested individuals with realistic punishments for defecting from those commitments. In looser, more open social networks, individualism or narrow self-interest (opportunism, free riding, etc.) is more likely to flourish, so that all must forgo many opportunities for mutual gain. Trust, and the norm of reciprocity associated with it, serve to reconcile self-interest and solidarity.

Although Forbes (61-3) contends that rational choice modeling can be both “‘positive’ and ‘theoretical’ without being ‘causal’ in the usual sense,” he nevertheless captures the exact sense by which I mean to suggest that rational choice modeling explains correlations:

Putnam's chain of correlations (from social networks through trust to democratic performance) gains its aura of causal necessity, not so much from the strength and persistence of the statistical relations he and others have able to demonstrate, as from the reasoning about collective action problems that accompanies the presentation of the still somewhat scanty evidence. (Forbes 2004, 60)

Similarly, Wantchekon (2004) uses rational choice modeling to explain a largely unrecognized regularity between civil war and democracy. According to his own empirical analysis, "nearly 40% of all civil wars that took place from 1945 to 1993 resulted in an improvement in the level of democracy." For instance, civil wars "gave birth to relatively stable democracies in Mozambique, El Salvador, Guatemala, and Nicaragua among others" (17). Wantchekon (17-8) notes that Immanuel Kant puzzled over how rational being with antagonistic interests could create order from chaos, and there has been no "systematic explanation of how a republican constitution can arise from an interaction between 'rational beings' and 'evil sentiments'." Wantchekon's game theoretic model proposes just such a solution.

Wantchekon's game (19-22) assumes that two warring factions have been involved in a costly and inconclusive conflict for control over a government. The key players of the game are two warring factions (and their citizen supporters) who compete for wealth and power, and a citizenry unaffiliated to either faction who are motivated primarily by concern for personal security. A stalemate in the civil conflict leads the two factions to simultaneously choose whether to invite a third outside party to negotiate a settlement (a Leviathan), to democratize, or to maintain the status quo. The sole source of wealth in

this society is the citizens' investment, and the warring factions' payoffs depend on the productive investment made by the residents of the country.⁷⁹ During the conflict, the warring parties expropriate wealth from citizens in order to bolster their respective causes.

The results of Wantchekon's game suggest that democratization is the most likely choice of the three available. The status quo not only ensures the continuation of insecurity for unaffiliated citizens, but also decreases incentives on the part of citizens to create wealth – because they fear wealth expropriation will only increase as the two stalemated parties ratchet up the conflict. Neither warring faction is likely to favour a continuation of the status quo in the face of a shrinking resource pie from which to expropriate wealth. The choice of a Leviathan is equally unappealing to all three parties. Unaffiliated citizens are unlikely to choose this option because: “[t]he tension between the desire to expropriate more (because of its effects on investment) is resolved in favor of more expropriation since Leviathan and its agent ‘compete’ over expropriation under the status quo.” The two warring parties are also unlikely to choose Leviathan, because Leviathan conceals its preference for which faction it will work with, and each faction knows that the “...faction that would not be favored by Leviathan would clearly be worse off than under the status quo and democracy.”

⁷⁹ When warring factions generate wealth from significant control of resources, drug distribution, or a foreign power, Wantchekon (19) notes that this assumption is violated. In these cases, democratization is much less likely to result.

Under the democratization option, all three stakeholders in the afflicted state stand to make gains. A neutral third party such as the United Nations oversees democratization, providing unaffiliated citizenry with enhanced personal security. Democratization also tends to increase citizen investments, because they no longer fear the same level of wealth expropriation from the now-disarmed competing factions. Moreover, wealth production increases on the part of citizens because competing factions in the democratic race for office are forced to offer lower tax rates.⁸⁰ And losing the election bears a lower cost than losing a military battle: “The faction that loses the election may be worse off because it loses its ability to expropriate, but its loss is moderated by the fact that citizens work harder and therefore generate a higher level of well-being for the whole society” (22).

Cederman (2003) provides a final example for how political scientists use mechanisms identified by rational choice modeling to explain regularities. Lewis F. Richardson’s *power law* tells us that the size of an event is inversely proportional to its frequency: there are few large earthquakes, but many small ones (135). Richardson found that the severity of interstate wars is power law distributed: for each ten-fold increase in the severity of wars, their frequency decreases by a little less than a factor of three (135). According to Cederman, Richardson’s power law is one of the most robust regularities in political science; yet it “is a regularity in search of a theory” (135). Cederman replicates Richardson’s empirical regularity but notes that “stronger confidence does not equal

⁸⁰ Wantchekon seems to assume that wealth production from lower taxation rates will ultimately generate more money for government coffers than will higher taxation rates.

conclusive corroboration, which requires considerably more accurate portrayal of the *causal mechanisms* generating the phenomenon in the first place” (145, my emphasis).

To that end, Cederman offers a sophisticated agent-based computer-generated model that mimics certain mechanisms found in physics, whereby tensions that build in a system periodically pass through thresholds unleashing disturbances of varying degrees.

Cederman models geopolitical change – most notably relating to technological developments – to see how “geopolitical instability changes strategic calculations by altering contextual conditions” (138). Technological changes present states with new opportunities that sometimes precipitate states’ decisions to expand territorially; and these decision, in turn, change the playing field for other actors in the system, which then make decisions based on the new geopolitical context. Cederman concludes that technological changes resemble bits of sand added to a sandpile: they add tension to a system that will often release itself with small-scale events and only rarely induces large-scale events.

Rational Decisions and Regularities

Unlike explanations identified by process-tracing, I contend that rational choice-based explanations are not based on regularities. This issue is important because critics of SR often contend that if mechanisms are comprised of regularities (as they are with process-tracing explanations), then the concept of mechanism is superfluous.

There *is* a large philosophical debate surrounding the nature of rational actions and their relation to regularities. Some contend that regularities govern the relationships between

beliefs and actions. On this view, propositional attitudes are separate and distinct from actions; beliefs and desires *cause* actions (e.g., see Davidson 2001), or else are inherently amenable to a D-N framework. Others, however, insist that beliefs and desires cannot be disentangled from actions; rather, the *reason* for an action is merely a “redescription” of that action (e.g., see Collingwood 1946, MacIntyre 1962, and Winch 1958). Although I would side with the former view, my argument does not necessarily hinge on this debate. The crucial point for my purposes is the *manner* by which rational actions generate regularities, regardless of whether or not any particular action is correlated with a given belief.

Consider, for instance, a *collective action* problem such as Hardin (1968) describes. Every single farmer in Hardin’s tale may add a herd to the pasture because they believe this will be beneficial to them. And it may be possible to subsume the connection between these beliefs and actions under a Hempelian-type covering law. But these factors will *not* help us understand *why* the environmental commons in question is inexorably degraded. This understanding comes only from *the strategic interaction* of individual rational decisions made by the commons’ farmers: each farmer’s decision to add a herd to the pasture is made based on calculations about whether other farmers with access to the pasture are likely to do the same. The unintended and undesirable consequence of environmental degradation does *not* stem from correlation-based or covering law-based knowledge; or, more accurately, this knowledge is incidental to the explanation for the inexorable environmental degradation. What *really* explains the environmental degradation – i.e., what does the “heavy lifting” in this explanation – is the

logic of collective action problems: the notion that when enough individuals in a collective act purely on selfish interests, public problems ensue.

iv. Conclusions: Intuitiveness and Acceptance

In this chapter, I have described two ontological properties of mechanisms – physicalness and agency – and two corresponding methods that generate findings that explain regularities in a unique fashion. Process-tracing generates findings that explain regularities as a chain of events or phenomena: this aspect of mechanisms is highly intuitive, and it is generally accepted in political science that process-tracing is in the business of causal analysis.⁸¹ Rational choice modeling generates findings that explain regularities as the aggregation or strategic interaction of individual choices. This type of explanation is somewhat less intuitive than the former, but although some might disagree (e.g., see Friedman 1996, Forbes 2004), political scientists often acknowledge that rational choice modeling *can* be in the business of making causal inference (e.g., see Bueno de Mesquita 1999).

The relationship between intentionality, interpretation, and regularities, explored in the following chapter, is not intuitive, and there is no widespread agreement in political science that interpretation can be in the business of causal analysis. Chapter VII then, is devoted to explicating how interpretation can generate findings that explain regularities,

⁸¹ This is not to suggest that political scientists generally agree about the exact *role* that process-tracing plays in making causal inference (there is strenuous disagreement on this point), only that most political scientists agree that process-tracing can play some type of role.

how it be in the business of making causal inference, and how some political scientists – especially IR *constructivists* – explicitly use interpretation to account for regularities.

Chapter VII: Intentionality

Whereas process-tracing identifies physicalness and generates findings that explain regularities as a chain of events or phenomena, and rational choice modeling identifies agency and generates findings that explain regularities as the aggregation or strategic interaction of individual choices, interpretation identifies intentionality and generates findings that explain regularities as being constituted by intersubjective meanings.⁶ In this chapter I explain and defend this latter proposition about the third ontological property of social mechanisms.

In the first section, I provide a brief sketch of the method of interpretation. In the second section, I address objections to merging interpretation and causal analysis, and explain how the method of interpretation can generate findings that explain regularities by closely examining and critiquing Charles Taylor's philosophy of interpretation. In the third section, I provide specific examples from *constructivist* literature in political science to support my argument for merging interpretation and causal analysis. The conclusions discuss compatibility and avenues for cross-fertilization between all three properties of mechanisms and their corresponding research methods. The conclusions also touch on the issue of how theoretical mechanisms are tested, although a full account of how this is achieved will have to wait until Chapter X.

⁶ As discussed further below, interpretation is intimately tied to what Weber called *Verstehen*: a researcher locates the social phenomenon within an intersubjective framework of shared meaning.

i. The Method of Interpretation

Interpretation holds that social phenomena – social practices, institutions, individual and group behaviour – are intrinsically meaningful and that their meanings are constituted by collective intentionality – shared desires, beliefs, norms, values, goals, etc. – or what interpretivists usually call intersubjective meanings. Social phenomena can be understood only by investigating the origins and nature of the intersubjective meanings of which they are constituted.⁸³ As Forbes (2004, 65) notes, intersubjective meanings are so *deeply enshrined* in the collective consciousness that they are “not normally topics for discussion or even reflection.” To “make sense of” – demystify, complete, clarify – social phenomena, the social scientist must determine the intersubjective meanings that underlie them. We make sense of social phenomena when there is coherence between the phenomena and its intersubjective meanings.⁸⁴

(Taylor 1994 [1971]) holds this latter view of what interpretation entails. In a similar vein, Geertz (1973, 1994 [1983]) contends that social research is conducted in much the same way as the literary critic analyzes a text. The goal is to demonstrate that actions and events have a deep symbolic meaning which is not immediately apparent: the focus is on showing that *something else* is going on besides what seems to be happening on the

⁸³ In this sense, linguistic analysis is sometimes central to interpretation. As we shall see, linguistic analysis is not incompatible with causal analysis.

⁸⁴ Some interpretivists use different terminology. Searle (1995), for instance, refers to intersubjective meanings as “collective intentionality.”

surface (Jones 1998, 40). For Geertz, humankind is inextricably caught up in a web of meaning that it has spun itself. Geertz (1994 [1983], 217) argues in favour of an ethnographic approach to the study of culture based on “thick description” and interpretation: “Doing an ethnography is like trying to read (in the sense of ‘construct a reading of’) a manuscript – foreign, faded, full of ellipses, incoherencies, suspicious emendations and tendentious commentaries...” “The culture of a people is an ensemble of texts, themselves ensembles, which the anthropologist strains to read over the shoulders of those to whom they properly belong,” contends Geertz (1973), in an effort to find out “just what the devil these people are up to.”

ii. Merging Interpretation and Causal Analysis

Some philosophers (e.g., Geertz 1994 [1983], Taylor 1994 [1971], and Winch 1958) insist that causal explanation and interpretation are mutually exclusive, or at least radically different activities. I argue against this position by analyzing and critiquing Taylor’s well known “Interpretation and the Sciences of Man.” I show that the distinction between interpretation and causation is not as sharp as Taylor claims.

The Case Against Merging Interpretation and Causation

Proponents of interpretation often contend that this method of investigation is directly opposed to the naturalistic methods that have dominated mainstream social science since the behavioural revolution of the 1950s and 1960s. These interpretivists contend that social scientists should abandon their effort to imitate the natural sciences. We will never

come to grips with the social world, proponents insist, by searching for objectively identifiable causal laws through inductive and deductive methods. Interpretivists sometimes point to the dearth of social laws promised by the behavioural revolution as proof that the naturalistic paradigm has failed.

Proponents of interpretation distinguish between explanation and understanding:

Explanation involves identifying causes of an event, whereas understanding involves discovering the meaning of an event or practice in a particular social context (Von Wright 1971, 5-6).⁸⁵ The proper focus of a social science, proponents maintain, is understanding through interpretation of meaning, not explanation through identification of causal laws.

Taylor (1985, 92), for instance, argues:

There is a constant temptation to take natural science theory as a model for the social theory: that is, to see theory as offering an account of underlying processes and mechanisms of society, and as providing the basis of a more effective planning of social life. But for all the superficial analogies, social theory can never really occupy this role. It is part of a significantly different activity.

Similarly, Geertz (1980) contends that social scientists should abandon their quest for objectivity and truth, and should likewise abandon paradigms of the natural sciences such as the discovery of causal relations, opting instead for other enterprises such as literary criticism and dramaturgy.

⁸⁵ This distinction should not be confused with the distinction that SR makes between explanation and prediction.

The Case for Merging Interpretation and Causation

From my perspective, interpretation and causal analysis are not necessarily antithetical. I contend that the dichotomies that interpretivists draw – between explanation and understanding, or identification of causal laws and interpretation of meaning – are based on a conflation of *means* and *ends*. Interpretivists and naturalistic social scientists often share similar ends, even if the means to achieve these ends are different ‘in kind.’

Note that it is not the case, as is sometimes maintained, that naturalistic social scientists confine their activities to identifying causal relationships. In fact, many natural social scientists are also interested in descriptions, definitions, and categorizations, or some combination thereof. By the same token, the activity of interpretivists is not confined simply to description. The ends of both interpretivists and naturalistic social scientists can sometimes be perfectly compatible. In order to demonstrate this, I will closely examine the philosophy of Charles Taylor.

As we saw above, interpretation, according to Taylor, aims to “make sense of an object of study.” But one might reasonably ask, “What exactly does it mean to “make sense” of a social phenomenon?” As an *end* this is fairly vague, or at least suggestive of various possibilities. I contend that, despite his (implied) objections to the contrary, it is not inconsistent with Taylor’s version of interpretation that “making sense” of a social phenomenon can include seeking out its cause or causes – that is, the reason or reasons why a social phenomena *exists* and *persists*.

Take Taylor's (1994 [1971], 186) discussion of "shame":

An emotion term like, "shame", for instance, essentially refers us to a certain kind of situation, the "shameful," or "humiliating," and a certain mode of response, that of hiding oneself, or covering up, or else "wiping out" the blot. That is, it is essential to this feeling's identification as shame that it be related to this situation and give rise to this type of disposition. But this situation in its turn can only be identified in relation to the feelings it provokes; and the disposition is to a goal that can similarly not be understood without reference to the feelings experienced...

In his critique of Taylor, Martin (1994, 264-5), points out that this passage "fairly bristles with causal concepts. Taylor assumes that the feeling of shame is brought about because of some humiliating situation, that is, it is caused by this situation; and this feeling brings about, that is, causes, a certain disposition to hide oneself, and so on." I agree with Martin that Taylor covertly smuggles in causal language. Note that in the above passage, Taylor contends that shame "gives rise to" a certain disposition to hide. As Martin's critique implies, "gives rise to" could easily be replaced with the word "causes." Try as he might, Taylor fails to expunge considerations of causality from the method of interpretation.

So far I have tried to establish that interpretation is not incompatible with the *end* of establishing causation. However, I want to emphasize that interpretation offers a unique *means* of explaining a causal relationship. Martin's critique of Taylor, above, establishes

“shame” as an intervening variable between the “humiliating situation” and the “action of hiding.” Martin contends that humiliating situations (e.g., an unreciprocated sexual advance) causes certain psychological reactions in the brain (e.g., shame), which in turn causes a reaction on the part of the individual (e.g., hiding, by trying to ‘lose’ themselves in the crowd). Thus Martin states, “When psychologists such as Freud expand and explain the insights of commonsense psychology, the connection of causality and shame becomes explicit.” Therefore, “shame”, for Martin, is a physical property of a social mechanism: it is part of a causal chain.

I want to suggest instead that interpretation explains seemingly connected observable events or a given correlation exactly as Taylor would describe: by “making sense” of the phenomena. Instead of thinking of “shame” as an intervening variable, we can think of it as a *background context* that “enables” a connection between the observable phenomena in question. As Taylor (1994 [1971], 187) notes, “We make sense of an action when there is a coherence between the action of the agent and the meaning of the situation for him.” Here, “shame” acts as a background context or intersubjective meaning that is “constitutive of” and therefore “makes possible” the connection between the human behaviour in question – the unreciprocated sexual advance followed by the actor’s attempt to ‘lose’ themselves in a crowd. As Taylor (1994 [1971], 195) argues, intersubjective meanings, such as that provided by “shame,” are “the background to social action.”⁸⁶

⁸⁶ Moreover, Taylor (1994 [1971], 186-7) points out that it is only against the context of “shame” that one can make sense of the type of “hiding” in question: “...the ‘hiding’ in question is one

To further my argument I now turn to the primary example from political science that Taylor draws upon in “Interpretation and the Sciences of Man”. In analyzing this example, I aim to demonstrate two related points: (1) that causation can be an ‘end’ for interpretation; and (2) interpretation provides a ‘means’ to that ‘end’ by identifying intersubjective meanings that “enable” the existence of regularities. The example from Taylor that I draw upon is his discussion of civilizational stability, or “social cohesion.” Taylor is concerned in this example with what accounts for social cohesion or a lack thereof in civilizations.

Although he studiously avoids the term, “cause,” Taylor’s concern is most easily understood precisely in causal terms. Consider, for instance, how Taylor (1994 [1971], 200) introduces this primary example:

It is an obvious fact, with which politics has been concerned since at least Plato, that some societies enjoy an easier, more spontaneous cohesion which relies less on the use of force than others. It has been an important question of political theory to understand what underlies this difference.

which will cover up my shame; it is not the same as hiding from an armed pursuer; we can only understand what is meant by ‘hiding’ here if we understand what kind of feeling and situation is being talked about.”

Can it be reasonably denied that Taylor is concerned with the following causal question: “What is/are the cause(s) of social cohesion or lack thereof?” The first point that I wish to make then, is fairly evident: the ‘end’ of interpretation can be causal in nature. We can “make sense” of social cohesion by understanding what causes it to succeed or fail.

Taylor (1994 [1971], 191) looks at the correlational linkages that mainstream political scientists have focused on with regard to social cohesion. In particular, he focuses on evidence from Seymour Lipset’s 1960 *Political Man*, which draws correlations between “incorporation of working class into the political process” (henceforth ‘integration’), “legitimacy,” and “social cohesion.”⁸⁷

It is fairly easy to imagine how integration and social cohesion might be connected causally by a series of intervening variables. For instance, when the working class is included in the political process of a nation, it gives them recourse (at least superficially so) to address their political and economic concerns. This, in turn, increases their confidence that their place in society is secured, which thereby enhances the legitimacy of the state. When a state enjoys legitimacy, social cohesion naturally follows: rules are

⁸⁷ Lipset (1960, 79-80) does not provide statistical-based correlations to connect these variables, although he does draw on observations that would support such correlations. (In other parts of *Political Man* (e.g., p.64), Lipset does draw on statistical-based correlations.) In any case, Taylor treats the connections between integration, legitimacy, and social cohesion, as connections that are based on correlational evidence. After all, Taylor (1994 [1971], 191) remarks earlier: “Political science as a body of knowledge is made up of such correlations...”

obeyed, laws are followed, and stability ensues. This may be especially true when a societal group such as the working class, which has high potential for destabilization, confers legitimacy. Thus, Taylor (1994 [1971], 202) remarks: “This approach that sees stability as partly the result of legitimacy beliefs, and these in turn as resulting partly from the way the status, welfare, access to political life of different groups fare, seems at first blush eminently sensible and well designed to help us understand the history of the last century or two.” But Taylor (1994 [1971], 202) contends that this type of correlational approach “has no place for a study of the intersubjective and common meanings which are constitutive of modern civilization. And we may doubt whether we can understand the cohesion of modern societies or their present crisis if we leave these out of account.”

Instead, Taylor looks to interpretation to “understand” social cohesion. One plausible intersubjective meaning that can help make sense of social cohesion, suggests Taylor, is the *society of work*. A society of work, contends Taylor, “is the ‘ideology’ that has frequently presided over the integration of the working class into industrial democracies...” By society of work, Taylor means, “the vision of society as a large-scale enterprise of production in which widely different functions are integrated into interdependence; a vision of society in which economic relations are considered as primary, as it is...with the tradition of classical utilitarianism...”; a vision in which “...there is a fundamental solidarity between all members of society that labor...for they

are all engaged in producing what is indispensable to life and happiness in far-reaching interdependence.”⁸⁸

I take Taylor’s explanation of how a society of work has “presided over” integration of the working class in industrialized democracies, as a cue for how interpretation can identify the intentional property of a social mechanism that explains the correlations between integration, legitimacy, and social cohesion. The entire causal connection can, on this account, be understood—that is, “made possible” by – the context of a society of work. The explanation thus runs as follows. A society of work has “presided over” the integration of the working class in industrialized societies, by providing them with an integral role in the interdependent society: to labour *is* to integrate oneself into the political process by fulfilling a vital role in the interdependent society. That is, to labour is to create solidarity in the “the great interdependent matrix of labour”; and hence to create legitimacy and social cohesion. Put otherwise, when the working class in modern industrialized societies produces their goods and services, they are fulfilling a crucial role in the society of work. As such, they are intrinsically engaged in the political process – whether they are cognizant of this fact or not. In Taylor’s terms, the society of work is “constitutive of” the integration of the working class – that is, it provides the vocabulary, not of words, but of ideas and actions, within which this integration is inexorably embedded. At the same time, their labour inherently legitimizes the state because it is founded upon the notion of a society of work. And so long as other sectors of society

⁸⁸ The plausibility of Taylor’s argument is incidental here: as Taylor himself notes, this argument is provisional and meant primarily to demonstrate the method of interpretation.

also fulfill their role in the “the great interdependent matrix of labour,” the production of goods and services by the working class ensures social cohesion: when everyone plays their role in society, harmony ensues.

In this sense, I contend that the society of work is an intersubjective meaning that possesses causal efficacy.⁸⁹ On my account, the society of work that characterizes modern industrialized nations is constitutive of the social phenomena in question – integration, legitimacy, and social cohesion. It is as if the spatio-temporal gap between these variables melts away when we consider the social context in which they are embedded. Interpretation, in sum, “makes possible” or “enables” the linkages between these social phenomena, thereby providing us with greater certainty that these connections are genuinely causal.

Skeptics might contend that my argument does not do justice to the method of interpretation. But, in fact, my argument does not stray very far at all from Taylor’s own position. For Taylor, intersubjective meanings do have explanatory power; that is, they help to explain social phenomena. Taylor (1994 [1971], 202) contends, for instance, that “the consideration of the granting of access to the political process [i.e. ‘integration’] as an independent variable may be misleading,” precisely because social phenomena that lack context have little explanatory value. To “make sense” of integration, contends

⁸⁹ “Causal efficacy” should not be confused with “causal effect.” We will see in Chapter XI that DSI equates causal effect with regularity. By “causal efficacy,” I mean the ability to produce a result.

Taylor, we need an intersubjective meaning such as the society of work. Thus Taylor (1994 [1971], 203) asserts: “My point is that it is certainly not implausible to say that it [the society of work] has some importance in explaining the integration of the working class in modern industrialized democratic society.”

But my argument does take Taylor a step further than he may be willing to go. I contend that intersubjective meanings do not merely have explanatory power for single social phenomena such as integration. They also have explanatory power for multiple social phenomena – such as integration, legitimacy, and social cohesion – that appear regularly connected. Intersubjective meanings provide an umbrella that unites these regularities: that is, it provides a permissive environment that “enables” these correlations.

Notice what is *not* being suggested by the interpretivist explanation offered above. The society of work does not play the role of intervening variable connecting integration to legitimacy, or legitimacy to social cohesion. There is no temporal process of disjointed events and phenomena here. Also, this explanation cannot be construed as a covering-law-explanation, for the relationship between integration and social cohesion cannot be construed as an ‘instance’ of the society of work. Nor does the society of work invoke the aggregation or strategic interaction of rational individual decisions: intentional social action is not based on a rational calculus but rather understood by analyzing the historical and cultural context of society (which includes rational actors). I do not deny the possibility of making these latter types of explanations, but interpretation identifies

intentionality rather than physicalness or agency, and so offers a unique type of causal explanation.

iii. Interpretation and Mechanisms in Political Science

The specific manner in which I suggest that interpretation and causal analysis can be merged may be somewhat unique, but the general notion of merging the two – although not without controversy – is not novel. In fact, it is an idea with notable philosophical lineage and is explicitly supported and championed by political scientists.

According to Max Weber's philosophy, interpretation and causation are not antithetical. "Sociology," claims Weber (1968, 4) "...is a science concerning itself with the interpretive understanding of social action and thereby with a causal explanation of its course and consequences."⁹⁰ For Weber, social science is distinct from natural science precisely because the social world is an inherently meaningful one. "Understanding," or what Weber termed, *Verstehen*, is intricately tied to interpretation; a researcher analyzes a social phenomenon from the vantage point of the actor(s) involved and locates the phenomenon within an intersubjective framework of shared meaning.⁹¹ Weber downplays the Hempelian models of explanation, arguing instead that empirical regularities must be subject to an interpretive analysis: "Statistical uniformities," writes Weber (1968, 100), "constitute understandable types of action in the sense of this

⁹⁰ Other philosophers of social science have taken up this Weberian perspective (e.g., see McLelland 1975, 89).

⁹¹ Weber considered both rational and irrational forms of meaningful action.

discussion, and thus constitute ‘sociological generalizations,’ only when they can be regarded as manifestations of the understandable subjective meaning of a course of social action.”

In political science, merging interpretation and causal analysis has become commonplace in the last decade in the fields of comparative politics (CP)⁹² and, especially, international relations (IR). Since roughly 1990, a growing tide of self-described *constructivists* – including prominent IR scholars such as Alexander Wendt (1999), John Ruggie (1998), and Peter Katzenstein (1996) – have attempted to forge a middle-ground between positivism and post-positivism. This brand of constructivism has become a major force in IR (Dessler 1999).⁹³

Constructivism’s core tenets are as follows: (1) international politics is explained at least as much by ideational factors as it is by material factors; (2) ideational factors, which include norms, conventions, and values, express *intersubjective* as well as subjective beliefs; (3) intersubjective meanings are not constant through time and space; (4) intersubjective meanings construct the preferences and identities of purposive actors,

⁹² Finnemore and Sikkink (2001, 404-5) note that the debate surrounding constructivism in comparative politics has been framed somewhat differently than it has in IR. Even if less explicit in CP than in IR, however, the debate does exist in the former. For an explicit discussion of constructivism and comparative politics, see Green (2002).

⁹³ In the last few years, constructivists in IR have begun to fine-tune their arguments, which is a testament to the maturity of constructivism in this discipline. See, for instance, Cederman and Daase (2003).

including individuals and states; (5) purposive actors and structures co-constitute one another in an ever-evolving process; and (6) identities and interests of states produce patterns of international outcomes (Copeland 2000, 187, Checkel 1998, 332, Finnemore and Sikkink 2001, 391, Ruggie 1998, 863, 879). Finnemore and Sikkink (2001, 393) sum up constructivism: “Constructivists focus on what Searle (1995) has called ‘social facts’ – things like money, sovereignty, and rights, which have no material reality but exist only because people collectively believe they exist and act accordingly. Understanding how social facts change and the ways these influence politics is the major concern of constructivist analysis.”

This epistemic community of political scientists argue *explicitly* that constructivism is perfectly compatible with causal analysis even as it holds onto methods of inquiry and theoretical beliefs that may not always be palatable to positivists (Copeland 2000, Checkel 1998, Dessler 1999, Finnemore and Sikkink 2001, Green 2002).⁹⁴ Yee (1996, 94) endorses “linguistic and interpretivist approaches to explain how the capacities of ideations enable mental events to produce their effects.” Experimental and quasi-experimental evidence must be “accompanied by a *causal story* indicating the mechanisms through which observed correlations evolve.” Yee (1996, 103) notes that

⁹⁴ Many other recent constructivist works (e.g., Coleman and Gabler 2002, or Windmaier 2004) are not explicit about their focus on causation, although they clearly offer causal explanations based in constructivism. This latter trend likely reflects a newfound *assumption* on the part of constructivists in IR that their analyses are inherently compatible with causal analysis. The battle on this epistemological front, in other words, has probably already been decisively fought.

although many interpretivists would reject his use of causal language, most are guilty of smuggling this causal language in surreptitiously. “[E]ven interpretive analyses of intersubjective meanings and discursive practices can be recast to offer some sort of causal or quasi-causal explanations. In his review of IR constructivism, Checkel (1998, 326) states, “It is important to note that constructivists do not reject the science of causal explanation.”

The brand of constructivism that these scholars champion – sometimes referred to as *conventional constructivism* (Hopf 1998) or *naturalistic constructivism* (Ruggie 1998) – shares certain starting points with its postmodern cousins – such as the centrality of intersubjective meanings – but, to paraphrase Checkel (1998, 325), IR scholars have rescued constructivism from the postmodernists.

Before I get more specific about the manner in which this epistemic community merges constructivism and causal analysis, it is important for purposes of my general argument in this chapter to clarify two related points. First, constructivism is essentially similar to interpretation. Second, constructivism, like interpretation and rational choice modeling, is a theoretically oriented *method*. Both of these points are important to my general argument because I want to show that interpretation/constructivism is a method that generates findings about intersubjective beliefs much as rational choice modeling is a method that generates findings about individual decisions and process-tracing is a method that generates findings about events and phenomena.

Constructivism: An Interpretivist Method

Although nearly all IR and CP constructivists recognize that there is a relationship between constructivism and interpretation, there is little agreement on the exact nature or strength of this relationship. Some constructivists (e.g., Finnemore and Sikkink 2001, 395) have attempted to distance themselves somewhat from interpretation. Others (e.g., Yee 1996) position their constructivist analysis explicitly within interpretation. I contend that constructivism is, at its core, indistinguishable from interpretation, although constructivists in IR have clustered around a particular set of intersubjective meanings: namely, national and transnational norms and values.

As demonstrated above, the focus of *both* interpretation and constructivism is, above all else, ideational factors, especially *intersubjective meanings* (often called *intersubjective beliefs* in IR and CP constructivist literature).⁹⁵ Nearly all constructivists recognize this fact. Ruggie (1998, 869), for instance, points out that constructivism “deals in the realm of ‘intersubjective beliefs,’ which cannot be reduced to the form ‘I believe that you believe that I believe,’ and so on.” Although it is true that constructivists have a unique focus on national and transnational norms and values, this difference belies the fact that these norms and values are, in fact, intersubjective meanings.

Constructivists offer several explanations for how constructivism differs from interpretation, but none are entirely accurate. Dessler (1999, 124) contrasts

⁹⁵ Fierke (2002) also attempts to link interpretation (especially in its postmodern form) with constructivism in that both share a focus on language.

constructivism with interpretation, arguing that the latter has “...no interest in the project of developing adequate causal accounts...” But this assertion is highly questionable, as the above discussion of Weber demonstrates.⁹⁶ Ruggie (1998, 857, 859) points out that Weber’s argument for the compatibility of *interpretation* and causation forms a core foundation for modern-day constructivism.

Other constructivists argue that constructivism constitutes just one of many streams of interpretation. Green (2002, 5), for instance, argues that interpretation includes postmodern, culturalist, feminist, and constructivist approaches. But this argument fails to account for feminist constructivist analyses (e.g., Allison 2001), and analyses that link post-modernism to IR constructivism (e.g., Fierke 2002).

Perhaps the most compelling reason advanced by some constructivists is that, in their estimation, constructivism is a *theory* whereas interpretation is a *method* (Katzenstein 1996b). Finnemore and Sikkink (2001, 392) contend that “Constructivism’s distinctiveness lies in its theoretical arguments, not its empirical research strategies.” Constructivism is often compared directly to two other major theories in political science, neorealism and *neoliberal institutionalism*. Several authors contend that constructivism is currently “no more” than a method, but strives to be a coherent theory. Checkel (1998, 325, 342), for instance, states, “[C]onstructivism remains a method more than anything else. The central challenge for these scholars is theory development.”

⁹⁶ It is worth noting on this point that Dessler (1999, 128) appears to generally underestimate interpretation, relegating its primary activity to mere description and categorization.

I contend, however, that the central contrast between constructivism on the one hand, and neorealism and liberal institutionalism on the other, is epistemological in nature. For instance, constructivists object to the methodological individualism and assumptions of rationality inherent in neorealism and neoliberal institutionalism.⁹⁷ This rejection has historical roots: the stage for Weber's rejection of methodological individualism is set by his espousal of interpretation (Ruggie 1998, 861). Neorealists and liberal institutionalists, for their part, eschew the causal force of intersubjective meanings. Moreover, despite the fact that constructivists have used a variety of methods (Finnemore and Sikkink 2001, 295) ideographic methods and thick description *remain* the hallmark of constructivist research.⁹⁸

In sum, I contend that, like rational choice, constructivism is a *theoretically oriented method*. Rational choice is theoretically oriented in the sense that its central focus is on

⁹⁷ Checkel (1998, 325) complains that IR constructivists have overemphasized structure to the neglect of agency. This neglect, claims Checkel, is somewhat mystifying given that constructivists generally acknowledge that structures and agents co-constitute one another. From my perspective, however, the focus of constructivism is rightly on intersubjective meanings and hence properly focused more on structure than agency. It is not that agents do not matter, but rather that, unlike, say rational choice modeling, their purposive action is not the ultimate, nor even the most important, components in the analysis of social phenomena.

⁹⁸ See, for instance, the articles in Katzenstein (1996). These articles make use of a variety of methods, but in each case they are supplemented heavily by ideographic methods and thick description.

individual-level human choices. Likewise, interpretation is theoretically oriented in the sense that its focus is on societal-level intersubjective meanings. Some IR and CP constructivists would agree. Ruggie (1998, 879) asserts: “[C]onstructivism is not itself a theory of international relations, the way balance-of-power theory is, for example, but a theoretically informed approach to the study of international relations.” And Green (2002, 9) argues that CP constructivism is marked by its “differing theoretical and methodological positions.”

How Interpretation/Constructivism Explains Regularities

I do not, by any stretch, wish to argue that constructivists are solely interested in generating explanatory mechanisms: constructivism is hardly monolithic. As Ruggie (1998, 867) points out, constructivists engage in non-causal empirical studies. Further, when constructivists do engage in causal analyses, the nature of these analyses varies. Checkel (1998, 334) notes that a major research strategy for many constructivists is to treat a norm as an *independent variable*, which is then correlated to the dependent variable in question. Eyre and Suchman (1996), for instance, explain Third World militarization as a consequence of a global institutional norm that legitimizes and encourages such militarization. They provide correlational evidence linking this norm to militarization in order to support their causal claim.

These caveats notwithstanding, I maintain that constructivism can identify the intentional property of mechanisms and generate findings that explain in the manner explicated above: namely, the mechanism acts a backdrop or context that “enables” the existence of

regularities. Thus, in Taylor's example discussed above, the society of work "makes possible" correlations between integration, legitimacy, and social cohesion.

Several constructivists hint at a similar approach to merging constructivism and causation. Finnemore and Sikkink (2001, 394) state:

Constructivists are skeptical about claims to all-encompassing truth [what Price & Reus-Smit (1998) call 'Big-T' claims] and instead produce and evaluate 'small-t' contingent claims. Such partial and contingent claims may still constitute causal explanation, albeit in a somewhat different sense than realists or liberals understand causality. For constructivists, understanding how things are put together and how they occur is not mere description...[A]n understanding of how sovereignty, human rights, laws of war, or bureaucracies are constituted socially allows us to hypothesize about their effects in world politics. Constitution in this sense is causal, since how things are put together makes possible, or even probable, certain kinds of political behavior and effects.

Yee (1996, 94-101) likewise focuses on the manner in which causal relations are made possible by certain contexts. By studying symbolic languages, a researcher can assess the causal relations of a particular actor or an institution. Language, notes Yee, "operates to define the range of possible utterances and hence the range of possible actions." In so doing, "language...authorizes or restricts, as well as prioritizes and distributes, the ideas and beliefs that policymakers can think and in so doing partly delimits the policies they can pursue." Language, in other words, gives rise to certain vocabularies because, as Taylor (quoted in Yee 1996, 98) has argued, "ideas do not properly exist before their

expression in language...That is what is meant by saying that language...is constitutive of thought.” Language not only defines the vocabulary within which discourse takes place, but also defines the “rules and conventions that govern the speech or utterances that are possible and hence in part the political actions that can pursue” (95). In this way, contends Yee (84-5), “[e]xplanation becomes a ‘two-tiered affair’, where statistical associations and [randomized experiments] generate valuable descriptive information that must be explained in terms of *causal* relations.”

Ruggie (1998, 869) shares a similar approach. Following Davidson (2001 [1963]), I reject Ruggie’s distinction between *reasons* for action and *causes* of actions.⁹⁹ But this distinction notwithstanding, the following passage reveals a similarity with my analysis of how constructivism explains:

Some ideational factors simply do not function causally in the same way as brute facts or the agentive role that [neorealists and neoliberal institutionalists] attribute to interests. As a result, the efficacy of such ideational factors is easily underestimated...Thus, the *aspiration* for a united Europe has not *caused* European integration as such, but it is the *reason* the causal factors (which presumably include bipolarity and economic interests) have had their special effect – in Weber’s words, produced an outcome that is historically *so* and not *otherwise*. (Ruggie 1998, 869, emphases in original)¹⁰⁰

⁹⁹ For more detailed arguments on the distinction between reasons and causes, see Collingwood (1946, 205-17), MacIntyre (1962), and Winch (1958).

¹⁰⁰ Like other interpretivists (e.g., Searle 1995), Ruggie (1998, 873) distinguishes regulative from constitutive rules, arguing that the latter provide “noncausal explanations.”

A Political Science Example

Price and Tannenwald (1996) use constructivism to explain a regularity relating to deterrence and weapons of mass destruction. Realists in IR have demonstrated in some empirical studies that states do not historically attack other states with chemical and nuclear weapons when they fear that such use would instigate a retaliatory act in kind (Waltz and Sagan 2002). Realists explain this regularity with the theory of *deterrence*, which argues that states, as rational, self-interested actors, are dissuaded from taking an action when doing so is likely to invoke unacceptable costs in return. Deterrence, note Price and Tannenwald (117) “provides the dominant explanation for the non-use of nuclear weapons by the superpowers during the Cold War and is often cited as the most important immediate factor in explaining the non-use of [chemical weapons].”

Price and Tannenwald, however, contend that although the realist theory of deterrence is not “entirely wrong,” it fails to adequately explain the non-use of chemical and nuclear weapons. An *adequate* explanation is lacking in two senses. First, deterrence theory fails to account for a number of historical episodes: in other words, there is notable unexplained variance in the empirical regularity linking conditions of deterrence (e.g., parity in capabilities) with the non-use of chemical and nuclear weapons. Second, deterrence theory fails to explain why chemical and nuclear weapons, as opposed to *other* types of weapons, produce situations of non-use. In order to remedy these failings, the authors aim to demonstrate “the socially constructed nature of deterrence and deterrent weapons” (114). “The patterns of non-use of these weapons,” write Price and Tannenwald

(115), “cannot be fully understood without taking into account the development of prohibitory norms that shaped these weapons as unacceptable ‘weapons of mass destruction.’”

I will focus on Price and Tannenwald’s argument with regard to chemical weapons in order to illustrate their points. Price and Tannenwald (119) note that deterrence theory fails to explain why, in the cases of the Spanish Civil War, the Korean War, the Vietnam War, and the Soviet intervention in Afghanistan, chemical weapons were not employed. In each of these historical episodes, chemical weapons could have been used effectively without fear of retaliation in kind. Moreover, deterrence theory does not explain why chemical weapons serve as a deterrent when other equally devastating conventional weapons do not: “How is it that a prohibitive fear of [chemical weapons] has operated over and above the fear of other powerful means of destruction, some of which are accorded the legitimacy of ‘conventional’ weapons despite their capability to wreak more havoc than [chemical weapons] can” (Price and Tannenwald, 122)?

Price and Tannenwald argue that to account for these empirical and theoretical failures a constructivist approach that “problematizes” the “variable of deterrence” is required:

Why would the fear of retaliatory [chemical weapons] attacks be any more robust a restraint than the fear of other horribly destructive methods of warfare such as incendiary bombing raids or submarine attacks on civilian shipping? If we are to avoid merely begging the question of why a special dread of retaliation operated with respect to [chemical weapons] we need to understand how the discursive practices of statesmen

served to set [chemical weapons] apart as a symbolic threshold of acute political importance and defined [chemical weapons] as a weapon *that might not be used*...In sum, the odium attached to [chemical weapons] is indispensable in accounting for their non-use. In the absence of a normative discourse that ostracized and politicized the use of [chemical weapons] as unacceptable, illegal, and reprehensible, a strong counterfactual case could be made for the possibility or even the probability that these weapons eventually would have been assimilated into military arsenals. (119-20, emphasis in original)

What accounts for the origin of the “chemical weapons taboo?” To answer this question, Price and Tannenwald trace the genealogical origins of this norm.¹⁰¹ “Rather than being viewed as simply the inevitable result of the objective qualities of chemical weapons, as is often supposed,” argue Price and Tannenwald (126-7), “the [chemical weapons] taboo is better understood as a political construction that owes much to a series of fortuitous events.” A measure to proscribe chemical weapons was accepted at the Hague Conference of 1899 “largely because it was not believed to have much significance,” for no such chemical weapons had yet been developed (127). Many subsequent efforts to ban chemical weapons were made on the basis of the Declaration. In the interwar period, the chemical industry launched “overzealous propaganda” campaigns “in order to secure chemical tariffs and the survival of the chemical warfare departments” (127). This campaign provoked equally overzealous campaigns on the part of opponents to chemical

¹⁰¹ Price and Tannenwald (124) make a distinction between a “genealogical method” and a “social constructivist method,” but nonetheless clearly categorize their overall analysis of chemical weapons as one that is social constructivist but that also employs a genealogical method.

weapons: “Hence the image of [chemical weapons] that was constructed was far out of proportion to the actual danger they represented...” (127).

These, and other historical events that emerged more by chance than by path dependency, created a norm that cast chemical weapons in an especially unfavourable light:

In genealogical fashion, an institutional tradition prohibiting [chemical weapons] came to be invoked as its own justification, in such a way as to obscure the fortuitous ancestry of the taboo. The [chemical weapons] taboo was reborn from the ashes of World War I not simply as a *technologically determined and self-interested* reaction to a prohibitively costly new means of warfare but also as a political construction whose institutionalization has in turn helped to politically legitimate the definition of [chemical weapons] as a practice beyond the pale of civilized nations. (129)

According to Price and Tannenwald, the norm surrounding chemical weapons that developed through a series of “fortuitous events” has had massive implications for international security – a significance lost on realist deterrence theorists.

In short, the “chemical weapons taboo” answers questions that realist deterrence theory cannot. Why there is unexplained variance in the regularity between conditions of deterrence and the non-use of chemical weapons: i.e., why were chemical weapons not deployed in situations where one party to a conflict had no reason to fear retaliation in kind? Why, in those instances in which deterrence theory does appear to explain the historical record, do *chemical* weapons – as opposed to equally devastating forms of

conventional weaponry – act as a deterrent? The answer, according to Price and Tannenwald, is that through a serendipitous combination of historical events, chemical weapons came to be seen as odious and unusable.

Price and Tannenwald's analysis point out that deterrence theory is incomplete and somewhat vacuous, but *not* wrong. Likewise the empirical regularity associated with the theory is weak, but *not* spurious. Generally, rationally acting and self-interested states *do not*, as the theory suggests, use chemical weapons when there is fear of retaliation in kind. But the theory *fails* to provide the *context* in which these “rational” decisions are made. *Rationality* is context-dependent: in the case of chemical weapons, rationality is tied inextricably to the taboo surrounding chemical weapons use.¹⁰²

Importantly for my purposes, Price and Tannenwald (145-148) engage in an epistemological discussion of the explanatory role of constructivism – one that supports my argument for how interpretation provides an explanatory mechanism for regularities. “Constructivism,” they argue “does not view the world in terms of discretely existing independent variables whose independent effects on variance can be measured according

¹⁰² Price and Tannenwald (125) distinguish between constructivist and rationalist approaches to norms: “In a rationalist view, norms constrain exogenously given self-interest and behavior or lead to recalculations of self-interest. In the constructivist view – developed primarily in the sociological literature – norms shape conceptualizations of interests through the social construction of identities.” In my view, this difference is not clear: as I argue below, constructivism and rational choice modeling can be compatible, as norms can influence preferences and interests.

the logic of statistics. Instead, certain issues, events, possibilities may matter – they become meaningful – only in the context of a norm.”¹⁰³

In other words, it is not that the “chemical weapons taboo” acts as an independent variable that can be measured statistically against the effects of deterrence. Rather, it is that deterrence must be viewed in the context of this taboo: chemical weapons deter *because* society has constructed a norm that makes their use morally deplorable and strategically questionable. As Price and Tannenwald (128, emphasis in original) contend, “*Norms structure possibilities, they do not determine outcomes.*”¹⁰⁴ “Constructivism’s contribution,” they note, “is that it evokes ‘context’ effects of norms” (150).¹⁰⁵

¹⁰³ Price and Tannenwald state the above in claiming that it makes little sense to attempt to test “norms” against “non-norms” arguments for the non-use of chemical weapons. But the particular quote that I employ is, as the authors state, a general point that “gets to the heart of a constructivist perspective on norms.”

¹⁰⁴ Price and Tannenwald are quick to point out that this explanatory role for norms – in which norms “possess the kind of taken-for-grantedness of a norm as intersubjective context” – is not the only possible one. Norms, for instance, can also “operate instrumentally,” as they do when they create public opinion constraints on decisionmakers.

¹⁰⁵ It might be tempting to attribute the explanatory role of the “chemical weapons taboo” to a covering law approach to explanation. Dessler (1999) notes that constructivist explanation can either be “particularizing” or “generalizing” – the latter being akin to a covering law approach. A comparison of this nature, however, is mistaken precisely because intersubjective meanings *provide* causal explanations whereas covering laws generally do not. As explained in Chapter II, covering laws instantiate a given regularity as a particular instance of a broader regularity but

iv. Conclusions: Cross-Fertilization and Testing Mechanisms

I have argued in this chapter and Chapter VI that mechanisms of social causation consist of three distinct ontological properties: physicalness, agency, and intentionality. Each of these three distinct ontological properties can be identified by three corresponding social scientific methods: namely, process-tracing, rational choice modeling, and interpretation. Each of these methods can generate findings that explain regularities in a unique fashion. Process-tracing generates findings on intervening chains of events and phenomena that link cause and effect. Rational choice modeling generates findings on the aggregation or strategic interaction of individuals' preferences that give rise to regularities. Interpretation generates findings on intersubjective meanings that are constitutive of regularities.

For analytical purposes I have presented these three properties of mechanisms as discrete, but this is not meant to exclude the potential for cross-fertilization between their corresponding methods: indeed, it may well be the case that identifying all three properties of social mechanisms will provide the most robust and compelling evidence available. Cross-fertilization between process-tracing and interpretation is one

generally fail to provide causal explanations for either the particular or the broad regularity. The explanatory context provided by intersubjective meanings, by contrast, provides this causal mechanism by making sense of the regularity. In the case at hand, the deterrent effect of chemical weapons is explained by reference to an historically developed norm that made their use morally reprehensible and strategically questionable.

possibility. As described above, for instance, Price and Tannenwald (1996) use process-tracing to demonstrate the series of “fortuitous” historical events that led to the creation of the “chemical weapons taboo” in international society. Their study suggests that cross-fertilization of process-tracing and interpretation can be useful.

Several IR and CP constructivists also point to the possibility of integrating rational choice modeling and constructivism. The basis for this integration is the notion that “rationality” is determined by social context. What appears to be irrational behaviour might in fact be a function of the societal context in which the action takes place.

Price and Tannenwald (1996, 151-2) make this point in relation to their study on chemical and nuclear weapons taboos:

Saying that the origins of the taboos are rationally based begs the question of what gets to count as rational and why. Once taboos of self-restraint exist, it may well be functional to uphold them (for either instrumental or constitutive reasons). But our question is the prior one of what constitutes “functional” or “rational.”...We are not giving a story about ‘irrationality,’ but one about what counts as rational. A rational deterrence argument is thus not necessarily incompatible with the more complex ‘taboo’ argument.

Constructivism asks a different set of questions and attempts to fill in the gaps that rationalist approaches leave unexplained...the goal is to convince deterrence and rational choice theorists of the incompleteness of their arguments, not to defeat them in some epic Lakatosian battle as if only can exist and the other must perish.

Similarly, Green (2002, 6) argues that, “After interpretivism, many argue that rationalism is trumped by the need to first determine preferences and goals, which are contextually and socially constructed.”

Finnemore and Sikkink (2002, 410-1) point to a set of IR literature that argues for a related avenue of cross-fertilization between constructivism and rational choice modeling. This set of IR literature, inspired by David Laitin, “argues that actors construct or choose their identities from a menu of existing choices. “The menu is historically and culturally constructed, but individuals choose rationally from the items that are on the menu at any given point.”

Not all IR constructivists are sanguine about the possibility of cross-fertilization. Ruggie (1998, 885), for instance, argue that rational choice and constructivism “are not additive, and they are unlikely to meet or merge on some happy middle ground.” But even Ruggie (1998, 885) acknowledges that “by pushing their respective limits in the direction of the other, we are more likely to discover precisely when one approach subsumes the other, when they represent competing explanations of the same phenomenon, when one complements or supplements the other, and when they simply describe different and incommensurate worlds.” Ruggie’s point might be extended to all forms of potential cross-fertilization between all three properties of social mechanisms and their corresponding methods. This potential is explored further in Chapter XIII.

Testing Theoretical Mechanisms

Three methods that can be used to identify three unique properties of social mechanisms were pointed out in Chapters VI and VII. But these methods provide only partial tests of the particular mechanisms that they identify. A more robust methodology is required to more fully test whether particular mechanisms operate as theorized. Price and Tannenwald (1996) use interpretation to generate the “chemical weapons taboo,” but (or so I will argue) further empirical testing is required to establish this theoretical mechanism with greater certainty. Chapter X explains how this additional empirical testing takes place with the SSR approach.

Before getting there, however, two further dimensions of mechanisms must be explored. In Chapter VIII, I explore the ontological status of social mechanisms: if social mechanisms consist primarily of *ideas*, on what basis can they generally be said to be literally true? And in Chapter IX, I demonstrate that theoretical mechanisms are always in want of deeper explanation.

Chapter VIII: The Reality of Ideas

Chapters VI and VII distinguished three distinct ontological properties of social mechanisms: namely, physicalness, agency, and intentionality. But I have yet to discuss the ontological status of these three properties. What exactly does it mean to say that mechanisms are *real* or *literally true*? This is a crucial question because SR is, at heart, an *ontological* position, which holds that unobservable entities and processes have a truth value: i.e., unobservables are literally true or false. If unobservable entities and processes are to have causal efficacy – that is, if they are to have a genuine role to play in causal explanations – then these unobservables must be real or literally true, for “what does not exist, cannot cause” (Miller 1987, 382).

This question is *especially* relevant for *social* scientific realism, because it is far *less* obvious for the social world than for the natural world what it even means to posit the “reality” of some thing in this world. This is because the social world is comprised primarily of *ideas*, which, by very definition, are *mind-dependent*. But SR holds that something cannot have a truth value unless it exists independent of the human mind. How can we say that a thing in this world, which is dependent on the mind, is “real” or “literally true”? This problem would seem to vitiate a *social* scientific realist approach to causation, for if ideas have no basis in reality, then ideas cannot have genuine causal efficacy.

The first section of this chapter discusses the main ontological claims of SR and the special problems these claims pose for SR in the *social* sciences. The second section utilizes and amends arguments made by Alexander Wendt (1999) in favour of SR. Following Wendt, I argue that although *ideas* are mind-dependent, they can nevertheless be said to have a truth value.

For purposes of my overall argument, it is critical to understand that I do not argue for the truth value of *particular* unobservables in this chapter: that is, I do not provide an avenue by which one can establish that a particular theoretical mechanism is literally true or false. The issue in this chapter is the *general* ground on which a case for the reality of social scientific mechanisms can be constructed, rather than whether and how one can reasonably claim any one *particular* unobservable mechanism is real or not. Put otherwise, the focus on this chapter is strictly ontological and does not concern the epistemological. Thus the “chemical weapons taboo” identified by Price and Tannenwald (see Chapter VII) can be said to have a basis in reality (because, if my arguments in the present chapter are accepted, social constructs in general have a basis in reality), but this particular intersubjective meaning requires further empirical testing in order to establish whether or not it is real.¹⁰⁶ The methodology for this latter task is left for Chapter X.¹⁰⁷

¹⁰⁶ Consider that even if SRs in the natural sciences have general arguments for why unobservables can be real, this does not mean that all unobservables are real. Empirical testing might prove or disprove the literal truth of any particular unobservable.

i. Scientific Realism's Ontology and its Application to the Social Sciences

Core SR claims center on the ontological status of unobservable entities and processes, such as subatomic particles and electromagnetic fields, postulated in scientific theories.¹⁰⁸

Most SRs make two key claims on this front: (1) Subject to the provisos that scientific methods are fallible and the knowledge they produce approximate and possibly misleading, scientific theories provide literally true descriptions of the unobservable entities and processes that they postulate¹⁰⁹; and (2) these unobservable entities and processes exist independent of the human mind (Chernoff 2002, 191-2, Brown 2001, 96-7).¹¹⁰

¹⁰⁷ A similar point applies to my discussion of epistemological issues in Chapters VI and VII.

These chapters point to methods that can be used to identify explanatory mechanisms, but these methods provide only partial tests of the mechanisms they identify. A more robust methodology is required to more fully test the epistemological status of theoretical mechanisms. As with ontological status, this issue of epistemological status is left to Chapter X.

¹⁰⁸ SR is not a monolithic philosophical position; numerous forms have been advanced, including, internal realism, critical realism, depth realism, and linguistic realism.

¹⁰⁹ SR is often taken to assert the *existence* of unobservable entities and processes. However, as Brown (2001, 97) points out, SR's claims are broader in that it asserts that theories have a "truth value." SR makes truth claims about the non-existence of things (such as the non-existence of God) just as it does about the existence of things (such as the existence of energy fields).

¹¹⁰ Unobservable entities sometimes become observable as technologies improve over time. Molecules, for instance, were unobservable until the advent of the electron microscope.

Notice that the second claim is, in a sense, necessary for and prior to the first claim. In order for scientific theories to describe reality, it is necessary that there be a reality to describe; and this reality cannot be one constructed in the human mind. Scientific theories do not create reality, but rather map a reality that already exists; or as Brown (2001, 97) puts it, “[truth is] something we discover (if we’re lucky), not something we invent.”¹¹¹

This second SR claim creates an important difficulty for social scientific realism. The social world, unlike the natural world, is comprised overwhelmingly of *ideas*. Whether or not one adheres to *idealism* – the notion that everything in this world is generated wholly by consciousness – most would acknowledge that ideas, by definition, are constructs of the mind. Hence, it is seemingly meaningless to say that social scientific theories provide literally true descriptions of the social world. If the social world were overwhelmingly mind-dependent, the external reality required by SR would appear not to exist outside the human mind.

What then does it mean to say that ideas have a truth value? For purposes of my overall argument, this is an important issue to clarify, most evidently in regard to the third ontological property of social mechanisms: intentionality. Recall that this property is suggestive of social constructions – ideas such as norms, conventions, and societal rules. Can the demands of SR for an external reality be reconciled with socially constructed ideas? Many constructivists would say no. Parker et al. (2003, 186), contend that

¹¹¹ This is sometimes discussed in philosophical literature as the subject-object distinction.

realism is at odds with interpretation precisely because interpretation requires an *idealist* ontology.

This issue also has bearing on the first and second ontological properties of social mechanisms. Physicalness does not simply refer to observable social entities. Case studies that employ the method of process-tracing routinely make reference to unobservable *social concepts*, such as ideology, poverty, power, social capital, political legitimacy, relative deprivation, and many others. Process-tracing routinely involves operationalizing these concepts¹¹², by finding adequate measures to represent them. GNP, for instance, is frequently employed to operationalize the concept of “wealth.” But measures of concepts rarely capture the entire concept. As Wendt (1999, 49) points out, “We might point to a speeding police car and say, ‘there goes the state,’ but that is the ‘the’ state, which consists of thousands of people, the structure of which cannot be seen.” To the extent that social concepts are simply ideas, it is important to state just how social SR can treat them as mind-independent, and therefore real.

With regard to agency, rationality is seemingly entirely mind-dependent. After all, rational decision-making processes take place in the minds of individuals. In what sense then, can SRs contend that rational decisions are real?

¹¹² Operationalization in process-tracing is similar to operationalization with statistical methods, although it amounts to a much less formal procedure.

ii. The External Reality of Ideas

In his bid to reconcile positivism, scientific realism, and constructivism, Wendt (1999, 47-91) addresses precisely the question that I raise above: namely, in what sense do ideas form an external reality? Is a “realism about ideas” incoherent? I employ Wendt’s three arguments for the external reality of ideas below, modifying his first argument and combining and modifying the latter two.¹¹³ I also contend that Wendt’s arguments are applicable to social concepts and rationality.

Wendt (71) points out that the notion that the social world is comprised largely of ideas “seems to violate the assumption of realism that the objects of science are mind/discourse-independent.” Skeptics maintain that “...even if science can know nature, it cannot know *society*,” for “...if social kinds are made of ideas then they do not exist independent of human beings” (49). Wendt (50) notes, “the issue is not settled even within the realist camp, with many realists about *natural* science arguing that the dependence of society on ideas makes a realist *social* science impossible.” Wendt (72-77) employs three arguments for the external reality of ideas, each of which “calls attention to ways in which social kinds remain objective despite their basis in shared ideas” (72). (Wendt focuses his remarks on social kinds rather than ideas per se, but as a constructivist he acknowledges that social kinds are ideational.)

¹¹³ Chernoff (2002) offers a powerful critique of Wendt, although he does not critique these particular arguments.

First, Wendt (72) contends that the social is underpinned at least in part by the natural. “Constructivism without nature,” argues Wendt, “goes too far.” Given that humans constitute ideas, Wendt contends that the physical structure of humans themselves demonstrates that the social is preceded by the natural:

In the case of physical artifacts, like ICBMs or garages, the material base consists in the physical properties without which these things cannot exist: a thing cannot be an ICBM if it cannot fly long distances, nor a garage if it is not big enough to fit a car...In the case of social kinds that involve people more directly, like states or professors, the material base consists in the genetically constituted properties of homo sapiens. Like other animals, human beings are *natural* kinds with certain intrinsic material properties like large brains, opposable thumbs, and a genetic predisposition to be socialized. Were it not for these material properties there could be no states or professors. Indeed, were it not for the materially grounded tendency of homo sapiens to designate things as “this” or “that”...there would be no social kinds at all. In the last analysis a theory of social kinds must refer to natural kinds, including human bodies and their physical behavior...

Wendt (73, 92-138) concedes that even if, contra most constructivists, it is not, “ideas all the way down,” the natural basis of ideas is relatively small. Notwithstanding this caveat, I believe that Wendt’s argument for the physical basis of ideas needs to be made somewhat more exact. In my view, what provides external reality to ideas is not simply that ideas are ultimately constructed by beings with a chemical and physiological makeup. Rather, what provides external reality to ideas is the notion that ideas might reduce ultimately to particular brain states or patterns. A scientist could plausibly point to a physical representation of person’s brain and attribute the state or pattern to a

particular idea.¹¹⁴ In fact, *brain imaging* is now a common and well-reputed science. Brain imaging uses Magnetic Resonance Imaging (MRI) technology and other technologies to capture actual brain states. If ideas can be reduced to particular brain states, then ideas can be said to have an external or objective reality: it makes the separation of subject and object possible.

Wendt's second and third argument for the external reality of ideas focuses on the extent to which ideas, *once constituted*, form a reality that "resist denials or misrepresentation" (73) and "confront the individual as social [fact]" (75). I believe that these two arguments, when combined and modified, help reconcile scientific realism's ontological commitment to a mind-independent world.

Wendt (75) points out that "Even though social kinds are not mind/discourse-independent of the collectivity that constitutes them, they *are* usually independent of the minds and discourse of the *individuals* who want to explain them." In other words, ideas/social kinds are created in the minds of a collectivity of individuals; but once so created, they exist regardless of the intentions or desires of the individuals that created them. Ideas, or social kinds, in this sense, "simply cannot be wished away."

Wendt (73) contends that part of what makes social kinds resistant to denials or misrepresentations is that they are "self organizing." Social kinds form from *within*, and

¹¹⁴ Freud believed that the inner-workings of the minds would eventually be found to be physiological (Skinner 1974, 11).

those external to the development of intersubjective meanings will be punished or rewarded according to how they respond to this reality.

As an example, Wendt invites us to consider the state, which as noted above, is an idea, the structure of which cannot simply be reduced to an income tax form or a police station. According to Wendt's argument, states are "self organizing" entities that depend only secondarily on external recognition:

The ability of a group to control and administer a territory (empirical sovereignty) historically has been the main consideration in its recognition by others as a state (juridical sovereignty)...A state's ability to organize itself as a state creates resistance to those who would deny its existence...Over time such resistance should bring others' theories about that state into line with its reality – i.e., resistance should lead to 'recognition' of its existence. The fact that a state is constituted by shared ideas does not make this resistance any less objective or real than the more strictly speaking material resistance of natural kinds. (73)

Wendt (74) acknowledges that external recognition plays a role in statehood. But "What makes, say, Germany, 'Germany' is primarily the agency and discourse of those who call themselves Germans, not the agency and discourse of outsiders." "The Spanish state was

a self-organized, objective fact for the Aztecs,” contends Wendt, “whether their discourse acknowledged this or not.”¹¹⁵

Popper (1965, 1979) deploys a similar argument to defend the reality of the ideational world. Popper (1979) contends that the world is divided into three ontological categories: *world 1* consists of physical bodies, such as rocks and planets; *world 2* consists of mental and psychological subjective states, such as pleasure and pain, and *world 3* consists of products of the human mind, such as theories and religious myths.¹¹⁶ *Materialist monists*, who admit only the corporeal, believe only the literal truth of world 1. *Dualists*, who admit the mental as well as the corporeal, believe in the literal truth of world 1 and world 2. But neither philosophical position, notes Popper (5-6), believes in the literal truth of world 3. But Popper insists that world 3 can be objectified. We know this because that which inhabits world 3 has a “*causal effect* upon physical things” (emphasis in original). Popper (11) explains further:

My fundamental argument in support of world 3 realism is very simple. We all know that we live in a physical world 1 which has been greatly changed by making use of science; that is to say, by using world 3 *conjectures of theories* as instruments of change.

¹¹⁵ Wendt (74-5) concedes that “social kinds lie on a spectrum of varying combinations of internal self-organization, and external social construction, the relative weights of which determine whether we should be realists or anti-realists about them.”

¹¹⁶ The tripartite ontology of social mechanisms that I advocate in this paper bears resemblance to Popper’s tripartite ontology of the world. Professor Fred Eidlin, Guelph University, first brought the relevance of Popper’s position to my attention.

Therefore, scientific conjectures or theories can exert a causal or an instrumental effect on physical things; far more so than, say, screwdrivers or scissors. (emphasis in original)

Popper (1965, 117) puts this belief in world 3 in a manner that is even more consistent with Wendt's claims: "Theories are our own inventions, our own ideas...this has been clearly seen by the idealist. But some of these theories of ours clash with reality; and when they do, we know that there is a reality...And this is why the realist is right."¹¹⁷

There is an important objection to the line of reasoning of Wendt (and hence, Popper and Searle) that, in my view, he unsuccessfully attempts to thwart. I offer a modification to Wendt's argument in order to defend his position.

Wendt (76) acknowledges that social kinds can change from within: "In effect, if a social kind can 'know itself' then it may be able to recall its human authorship, transcend the subject-object distinction, and create new social kinds." Wendt points out that "such

¹¹⁷ Searle (1998, 111-134) defends the objective reality of the ideation world in a similar fashion to both Wendt and Popper. "...[H]ow can there be an objective social and institutional reality that is the reality it is only because we think it is?" asks Searle. "The answer is that the collective assignment of status functions, and above all their continued recognition and acceptance over long periods of time, can create and maintain a reality of governments, money, nation-states, languages, ownership of private property, universities, political parties, and a thousand other such institutions that can seem as epistemically objective as geology and as much a permanent part of our landscape as rock formations" (131-2).

reflexive potential is inherent to social life...,” but could vitiate the argument for the external reality of ideas since “reality is being caused by theory rather than vice-versa.”

Wendt’s response to this challenge is to suggest that such reflexivity is relatively uncommon. “If societies were constantly...in a sort of ‘permanent conceptual revolution’,” contends Wendt, “we could not be realists about society.” Fortunately for his theory, Wendt seems to conclude (without much evidence) that social flux is rare.

I am less certain than Wendt that reflexivity of social kinds and intersubjective meanings is rare. Consider, for example, how intersubjective meanings surrounding race, gender, human rights, and environmental norms, have changed over the course of a few decades. One would be hard pressed to argue that gender today is the same social kind as gender in the 1950s. Moreover, as Searle (1998, 132) points out, the “withdrawal of collective acceptance,” can have dramatic consequences, “as witness the amazing collapse of the Soviet empire in a matter of months, beginning in *annus mirabilis* 1989.”

Fortunately, I do not believe that social kinds must be immutable in order to be real. Even if change comes from within the same collectivity that initiates the intersubjective meaning – that is, if the change is truly reflexive – reality is conferred by the fact that these intersubjective meanings do, even if for a brief period of time, constrain individual behaviour. Put otherwise, processes of reflexivity or recursivity can be objectified: there is always a point in the loop when social kinds are objective. Western conceptions of gender in the 1950s constrained behavior in various ways. These Western conceptions of

gender have changed significantly since the 1950s, but this change does not mean that gender now fails to constrain behaviour: rather, the types of constraints have changed.¹¹⁸

iii. Conclusions: Application to the Three Ontological Properties of Mechanisms

Following Wendt, I have argued that ideas have a truth value, based foremost on the notion that ideas, once formulated, take on a life of their own. This argument is most obviously applicable to intersubjective meanings (which relate to intentionality, the third ontological property of social mechanism) and social concepts (which relate to physicalness, the first ontological property of social mechanism). Social norms and concepts might originate in the minds of individuals; but once norms take hold in society, they constrain individual behaviour.

The argument is equally valid for rational decisions (which relate to agency, the second ontological property of social mechanisms). As argued in Chapter VI, it is the *aggregation* or *interaction* of rational decisions that produces certain social outcomes (e.g., Hardin's "tragedy of the commons," or "prisoner's dilemma"). As with

¹¹⁸ It may be objected that my argument (and hence, Wendt's argument) is circular since it appears to rely on causal efficacy to demonstrate reality, and causal efficacy is not possible without reality. However, this objection is not quite accurate. The demonstration of "causal efficacy" is better thought of as the demonstration of just "efficacy" or "impact." If one is able to demonstrate that an intersubjective meaning constrains individual behaviour then one can infer the reality of that intersubjective meaning; but one does not infer efficacy from reality: rather, one infers efficacy from an empirically demonstrated relationship between an intersubjective meaning and individual behaviour.

intersubjective meanings and social concepts, rational decisions ultimately produce outcomes that defy denial or misrepresentation. The rational decision-making process of any given individual may be entirely inside the mind, but in combination these decision-making processes develop an (often unintended) external reality.¹¹⁹

¹¹⁹ Friedman (1996, 2) points out that economists often simply assume rational choice's postulates to be obviously true, placing the burden on skeptics to prove otherwise. The notion, developed in this chapter, that rationality can create an external reality at least provides a basis for the literal truth of these postulates. But in order to know if any *particular* rational choice-based mechanism is working as hypothesized, further empirical testing is required: as Friedman (3) quoting Buchanan and Tullock argues, "the ultimate defense of the economic-individualist behavioural assumption must be empirical..." The methodology for this further testing is discussed in Chapter X.

Chapter IX: Mechanisms and The End of Explanation

Chapters VI, VII, and VIII, developed epistemological and ontological dimensions of social mechanisms. In Chapters VI and VII, I argued that social mechanisms are comprised of three ontological properties: physicalness, agency, and intentionality. Process-tracing, rational choice modeling, and interpretation, respectively, identify these properties and generate findings that explain regularities in a unique fashion. In Chapter VIII, I argued that social mechanisms could be said to be real or literally true despite the fact that the social world is comprised of mind-dependent ideas.

But although process-tracing, rational choice modeling, and interpretation, go part way toward testing the theoretical mechanisms that they identify, more robust testing is required to demonstrate that they operate as theorized. And although it is one thing to establish a *general* basis by which it makes sense to talk about the reality of unobservables in the social world, it is quite another to establish the reality of a *particular* theoretical mechanism. The next step forward in the development of the SSR approach to causation then, is to argue for a methodology that can establish social mechanisms with greater certainty. However, before taking this step forward in Chapter X, it is necessary to first take a step to the side in this chapter to deal with another dimension of mechanisms that has bearing on the relationship between mechanisms, regularities, and causation.

The argument for mechanisms requires an important caveat. Explicitly or implicitly, SRs sometimes portray mechanisms as the *end of explanation*. On this account, a theoretical mechanism ‘fills in’ the black box in its entirety: a mechanism represents a *seamless process* between cause and effect, and in itself is in no further need of explanation.¹²⁰

The iconic SR, Wesley Salmon, whose work has inspired SRs in both the natural and social sciences, explicitly holds this view of mechanisms. For Salmon, mechanisms represent a seamless process connecting cause and effect: “not a *chain* of causation”, as Salmon (1998, 205, 301) puts it, “but rather a *rope* of causation.” SRs in the social sciences sometimes imply a view consistent with Salmon.¹²¹ For instance, Dessler (1991) treats the process of adiabatic lifting as a seamless process (see my discussion of Dessler in Chapter X). Thagard (1999) talks about a “continuous process connecting infection and the development of ulcers.”

If a mechanism represents a seamless process, the quest for causation is, in some sense, finite. A certain level of uncertainty always exists – the explanatory mechanism is always subject to falsification – but further explanation of the mechanism is not required. In this chapter I argue, conversely, that every mechanism itself demands explanation. It

¹²⁰ The idea for this line of argumentation came, in part, from a private communication with University of Toronto Professor of Philosophy, Danny Goldstick, October 2nd, 2001.

¹²¹ This is sometimes the case, but not always. Little’s (1991) “causal mechanism thesis,” for instance, clearly favours the idea of a “chain of causation” and not a “rope of causation.” Bennett and George (1997, 7) contend that “each step or link [in] a causal process should be supported by an appropriate ‘law,’ defined for historical explanation by Carl Hempel as a statement of a regularity between a set of events.”

is not possible to construct a seamless process between cause and effect because there is no end to explanation. The quest for causation, in this sense, is infinite. In the following chapter I argue that, for pragmatic purposes, the quest for causation is usually terminated at a certain point: epistemic thresholds exist beyond which deeper explanation is not required to make causal inference. But my argument in this chapter is purely theoretical: causation involves a never-ending process of deeper explanation.

It is important to note at the outset a possible shortcoming to my argument against an end to explanation. The phenomenon of quantum indeterminism suggests that there may in fact be an end to explanation.¹²² In making this exact point, Glennan (1996, 66) notes: “Sooner or later the process of decomposition of a system into parts must come to an end.”¹²³ This caveat notwithstanding, it is questionable as to whether *all* physical

¹²² Glennan (1996, 66-7) provides an example of quantum indeterminism. “It is possible to construct a device [that] shoots a pair of particles in opposite directions to distant targets. These particles can be prepared in such a way that, upon hitting these targets they will deflect in one of two directions, up or down. It is not possible using this preparation technique to determine in advance which direction they will go. However, quantum mechanics predicts and experiment confirms that if one particle deflects up then the other particle will deflect down and vice versa.” Even when all conceivable sources that might account for this result are controlled (e.g., the possibility that the two particles are sending signals to one another), physicists cannot predict which particle will go up and which will go down, only that they will go in opposite directions.

¹²³ Glennan (1996) attempts to save his “mechanical theory of causation” – which holds that causation is decomposable into successively lower layers of mechanisms – by arguing that there is a demarcation between causal and non-causal levels of explanation. He argues that quantum

processes are ultimately decomposable into physical quantum effects; and it is even more questionable as to whether all social processes are decomposable into physical (and, ultimately, quantum) processes.

It is important to note at the outset as well that in this chapter I argue that a regularity can often be decomposed into a chain of lower-level regularities; and each link in this chain of regularities is then itself decomposable into a chain of even lower-level regularities, and so on in a near-infinite regress. This argument raises issues with regard to the relevance of the term mechanism. It is common to refer to each of these causal chains as a “mechanism” (e.g., see Bennett 1997, Glennan 1996, Little 1991). But if causal chains are composed of nothing more than regularities, is “mechanism” a genuinely relevant concept? This issue is taken up in the following chapter.

The first section of this chapter examines Salmon’s argument in favour of a “rope of causation.” I contend that this argument is faulty, casting doubt on the SR argument in favour of an end to explanation. The second section addresses an important argument made by Jerrold Aronson that could derail my argument against Salmon. The concluding section relates my argument against an end to explanation to the three ontological properties of social mechanisms, and more generally, to the social scientific quest for causation.

indeterminism occupies the level of non-causal and therefore does not figure in actual causal explanation.

i. Salmon and the End of Explanation

In Salmon's first version of physical causation, his mark transmission theory, an entity continuously propagates its internal structure – and any alteration to its internal structure – thereby transmitting its influence. In his second version of physical causation, his conserved quantities theory, a single causal interaction involves an exchange of physical quantities such as momentum, thereby transmitting its influence. (See Chapter III for more details on Salmon's two theories of causation.) Salmon (1998, 205) argues that his notion of a *fluid* causal process flies in the face of the Humean notion of a causal chain of *disjointed* events:

Hume and most other philosophers who have discussed causality have thought of a cause and its effects as distinct events. If they are separated in space, or time, or both, one tries to find intermediate events that provide a connection between them. By this means one tries to construct a causal chain...I would advise thinking of a thread or a cord instead of a chain. They are continuous; they are not composed of links. Thus, there is not question about the power to produce the next event because no such next event exists. In a causal process, the causal influence is transmitted continuously.

Does Salmon's (1998, 204) bold claim to solve "Hume's problem of causation" stand up? Does the mechanism posited by physical causation represent a seamless process? I argue that Salmon's arguments fail: the supposed "rope of causation" rests on fundamental laws of physics, which themselves require ever-deeper explanation.

Consider Salmon's conserved quantities theory. In this physical theory of causation, an exchange of conserved quantities represents the mechanism between cause and effect. Salmon (1998, 253) claims: "...causal processes transmit conserved quantities; and by virtue of this fact, they are causal." In his "broken glass" example (see Chapter III), a baseball and a molecule of nitrogen strike a glass window simultaneously, causing the window to shatter. How do we know that it was the baseball and not the nitrogen molecule that caused the window to shatter? Because, the momentum of the incoming nitrogen molecule is vastly smaller than the momentum of the outgoing shards of glass as they fan out into the air. The momentum of the incoming baseball, conversely, is roughly equal to that of the outgoing shards of glass. The causal interaction of the baseball and the window involve a transmission of conserved quantities (i.e., a mechanism); so does the causal interaction of the nitrogen molecule and the window, but this latter interaction is minute in comparison.

Salmon's formulation of a causal mechanism is compelling, but he seems to overlook the fact that his theory of mechanism cannot escape regularities. More specifically, the theory appeals to the laws of conservation, which include the elementary laws of conservation of momentum and conservation of energy. Although these laws are considered to be 'elemental', they do not represent an end to explanation. Rather, these laws represent regularities that are themselves in need of explanation by deeper principles. For instance, the Standard Model of particle physics explains (among other

things) the laws of conservation of momentum.¹²⁴ Thus, Salmon's mechanism that purports to represent a seamless process (and thereby represent genuine causation) rests on a regularity that itself requires explanation.¹²⁵

This inextricable link between regularities and mechanisms is common to most scientific inquiries. Take, for instance, the gas laws of Boyle and Charles, which are empirical regularities. Boyle's law states that as the volume of a gas rises, so does its pressure. Charles' law states that as the temperature of a gas rises, so does its pressure. These gas laws are considered 'elemental' in the field of chemistry, but they cannot be said to represent an end to explanation because they are each explained by the kinetic theory of gas. The kinetic theory of gas describes the microscopic behaviour of molecules and the interactions that lead to macroscopic relationships like Boyle's and Charles' laws. Does the kinetic theory of gas, then, represent a seamless process between cause and effect? It does not. The kinetic theory of gas, in fact, relies heavily on fundamental physical laws. For instance, a key assumption of the kinetic theory of gas is that molecules observe Newton's Laws of Motion. Again, we see that mechanisms do not represent a seamless process because there can be no 'end of explanation.'

¹²⁴ The Standard Model of particle physics describes the strong, weak, and electromagnetic forces, as well as fundamental particles that make up matter.

¹²⁵ Dowe (2000, 95), whose own conserved quantities theory of causation inspired Salmon's version, seems to worry about the issue that I raise here: "...regularities are not by any means the only form of evidence about conservation laws," he argues, "theoretical considerations are also important."

In fact, my case can be strengthened, because in order to make a causal claim using Salmon's conserved quantities theory, it may not simply be enough to draw on the laws of conservation. Other laws, as well, may play an important role. I will use Salmon's own "broken glass" example, discussed above, to make this point.

The law of conservation of momentum is premised on an important assumption: namely, it is assumed that momentum is conserved in systems that are *isolated*. Truly isolated systems are not possible in the physical world. Nevertheless, many interactions between moving objects can, for all intents and purposes, be treated as isolated systems because the external factors acting on the interaction are minimal. For example, consider two hockey pucks that slide along a sheet of ice and collide with one another. The individual motions are influenced in a small way by gravitational forces of immediately surrounding objects such as the walls of the rink, and to an even greater extent by friction with the ice surface and surrounding air. But the interaction between the two pucks so overwhelms these forces they can usually be discounted.

In Salmon's example, we are asked to assume that the system in question – that of the nitrogen molecule and the baseball hitting the windowpane – can be treated as an isolated system. Salmon notes that the momentums of the incoming baseball and the outgoing shards of glass will be "roughly" equal. Further, we are to assume that the *degree of similarity* between the momentum of the incoming baseball and that of the outgoing shards of glass *is so much greater* than the degree of (dis)similarity between the

momentum of the incoming nitrogen molecule and that of the outgoing shards of glass, that it is obvious as to which is responsible for the causing the event in question.

Now, this latter assumption may be a safe one when comparing a nitrogen molecule to a baseball. But what happens if we were to compare, say, a small coconut to a baseball? Suppose that a baseball strikes a pane of glass a fraction of a second prior to the coconut and we have no way of determining which of the two actually struck first. In this case, the momentum of the incoming coconut *may also be 'roughly equal'* to that of the outgoing shards of glass. In order to actually determine causation in this case we would require knowledge of the external forces (e.g., gravity or friction) acting on the systems. These external forces might just make up the difference in momentum between the two objects in question. To determine causation then, would require fairly precise calculations regarding these external forces, which in turn would necessitate the need for laws relating to these forces. Thus, Salmon's theory might embody other laws in addition to the laws of conservation of momentum.¹²⁶

ii. Aronson and the End of Explanation

It might be argued that Salmon's conserved quantities theory is an *ontological* theory of causation *not* an *epistemological* one – a distinction that I fail to make. That is, Salmon purports to show exactly *what* happens in the event of a physical interaction between two

¹²⁶ By his own account, Salmon (1998, 252, Salmon 1984, 148) abandoned his first theory of causation because of its reliance on counterfactuals. I have argued here that his second theory of causation is similarly flawed because of its reliance on regularities.

processes. He does *not*, however, aim to show *how we know* what happens. The distinction between ontology and epistemology could be crucial because a purely ontological theory of causation need not concern itself with the issue of whether or not fundamental laws are drawn upon. That is, it could be argued that there is simply no need to draw on fundamental physical laws; these laws are *not* at the heart of the theory. Instead, the heart of the theory is the *very exchange of momentum or energy*. This would mean that an end to explanation could merely be assumed, and thus a seamless process legitimately constructed.

This is precisely the defence employed by Jerrold Aronson – the originator of Salmon’s second version of physical causation. Aronson (1982, 295, 301-2) argues that momentum or energy would be “transferred” *regardless of whether laws of conservation existed at all*: “...[I]t is the process of transference,” he contends, “which makes the sequence causal...” Aronson notes that this feature of his *transference theory of causation* is a natural outgrowth of its *ontological*, rather than *epistemological* basis: “My concern here is different from Hume’s: his was with the epistemological basis for a causal relation, while mine is with the ontological nature of the causal relation as an object of scientific thought” (295).

Unfortunately, there is a difficulty with Aronson’s argument. Consider the example he uses to illustrate his theory:

For the purposes of illustration, consider the following experiment. It consists of two spheres, *a* and *b*, one moving and the other at rest at time₁. At time₂, *a* makes contact with

b, [and] comes to a halt while *b* moves off at time₃. In other words, *a* lost all of its momentum in this sequence of events while *b* gained momentum. (293-4)

Aronson (194-5) points out that in this example we can refuse to accept that *a* caused *b* to move only if we are prepared to accept a most unlikely coincidence:

In the case of *a* causing *b* to move, *b*'s momentum at time₃ is the same as *a*'s at time₁. Denying this would lead to quite a mystery, for then we would have to say that somehow *a* lost all of its momentum at time₂ and *b* acquired the exact same amount at that very moment!

He concludes: "It would be more natural to regard *a*'s loss and *b*'s gain as stages of a single process of transference of one and the same quantity rather than, á la Hume, separate and distinct events" (295).

But Aronson's example does not live up to the promise of his theory. The example begs the question of how we could possibly know in the first place that it is an exchange of momentum that causes sphere *b* to move when hit by *a*. Aronson's answer seems to be that this is obviated by the fact that sphere *a* comes to a complete halt at the very same moment at which sphere *b* takes off with equal momentum. But suppose for a moment that sphere *a* was significantly smaller in mass than sphere *b*. In this case, momentum may indeed be transferred in the collision, but sphere *b* may only move an imperceptible amount. This shows that our knowledge regarding the transference of momentum cannot be reduced to the observation of a single collision between two spheres.

If Aronson were being more candid, he might concede that the basis for his theory of causation rests on the well-established laws of conservation of momentum. This means that the ontological cannot be successfully separated from the epistemological: *what* we know depends on *how* we know it. Acknowledging this much rules out any end to explanation that is founded on ‘bare’ causal facts: If we appeal to laws – in this case the laws of conservation of momentum – then we will always need to further our explanation of these laws (hence the development of the Standard Model). Consequently, the notion of a seamless process is ruled out.¹²⁷

Conclusions: Social Mechanisms and The End of Explanation

The arguments developed in this chapter relate most evidently to the first ontological property of mechanisms: physicalness. It suggests that regularities are decomposable into a chain of lower-level regularities; and each link in this chain is then itself explained by a chain of even lower-level regularities, and so on in a near-infinite regress. As noted at the outset of this chapter, this argument raises an issue with regard to the relevance of the term mechanism, which is taken up in the following chapter.

For immediate purposes, however, it is important to note that my argument against the end of explanation applies equally to the second and third ontological properties of social mechanisms: agency and intentionality. As argued in Chapters VI and VII, unlike process-tracing, rational choice modeling and interpretation generate findings that do not

¹²⁷ Again, quantum indeterminism might vitiate my argument.

reduce to regularities. Nevertheless, the findings generated by rational choice modeling and interpretation can always be explained in terms of deeper principles.

Consider first rational choice modeling. We will see in the following chapter that SRs and IEs take opposing positions on important assumptions regarding rational choice. Generally, IEs assume that all humans make choices according to uniform principles: they champion so-called “thick rationality.” SRs, conversely, contend that human choice is context dependent: they champion so-called “thin rationality.” Neither uniform rational choices (thick rationality) nor context-dependent rational choices (thin rationality) represent an end to explanation. Context-dependent rational choices (e.g., the notion that humans sometimes make decisions on the basis of non-self-interested factors) clearly involve explaining particular contexts (e.g., In what cultures might we expect to find non-self-interested factors as bases for decisions?). Uniform rational choices involve untheorized starting assumptions. But untheorized assumptions are not satisfactory. The fundamental postulates of thick rationality (e.g., the notion that humans maximize their utility preferences) can always be explained in terms of deeper principles (e.g., Are there evolutionary advantages to maximizing utility preferences?) (see Lalman et al. 1991, 97, Little 1991, 62-5, and Frochlic and Oppenheimer 1984). In brief, explanation does not end with an examination of agency.

The same general argument applies to interpretation. Intersubjective meanings do not represent an end to explanation. One might question, for example, what deeper principles give rise to Taylor’s “society of work.” What is it about Western civilization that

generates “solidarity between all members of society that labor?” What historical forces have given rise to this phenomenon? Thus, as with physicalness and agency, explanation does not end with an analysis of intentionality.

In sum, this chapter has argued that all three ontological properties of mechanisms are always in want of further explanation. But this argument in this chapter raises an important issue: If there is no end to the depth of theoretical mechanisms, is causation a relevant concept? How does one know when an explanation has gone deep enough to be considered “causal.” These questions are taken up in Chapter X. The argument for purposes of this chapter is purely theoretical: SRs who contend that the quest for causation is finite because mechanisms represent seamless processes between causes and effects are mistaken. Save for the possibility that social mechanisms end at the quantum level, the quest for causation is infinite.

Chapter X: The Epistemological and Ontological Status of Mechanisms

This chapter ties together various strands from the arguments thus far in Part II, and contains the final developments of my SSR approach to causation. Chapters VI, VII, and VIII, developed the epistemological and ontological dimensions of mechanisms. I argued for three properties of theoretical mechanisms, each of which is identified by a particular social scientific method, and each of which can be said to have a truth value despite being comprised of mind-dependent ideas. But although process-tracing, rational choice modeling, and interpretation, generate findings that go *part* way toward testing particular theoretical mechanisms, more robust testing is required to ensure that the mechanisms operate as theorized. And although it is one thing to establish a *general* basis for the reality of unobservables in the social world, it is quite another to establish the reality of *particular* theoretical mechanisms. In this chapter I provide a methodological approach – the argument from coincidence (AfC) – by which the epistemological and ontological status of theoretical mechanisms can be more firmly established.

In order to understand the role that the AfC plays in the SSR approach to causation, it is necessary to refine the debate over causation. Part I of this thesis pitted Humeans against SRs, and indeed that is how the debate over causation is often portrayed. But based on differing approaches to the epistemological and ontological status of mechanisms and the ontological status of unobservables, I argue that other positions on causation are possible and require attention. In fact, we will see in Part III of this thesis that in the discipline of

political science, most of the tension over causation is *not*, as many SRs claim, between Humeans and SRs.

In addition to the above agenda this chapter also ties up two loose strands from previous chapters in Part II. First, I explain why mechanism is a relevant concept when – as is the case with physicalness, the first ontological property of social mechanisms – the explanation of a regularity consists of nothing more than a series of regularities. Second, I explain why causation is a relevant concept when – as I argued in Chapter IX – there is no end to explanation and hence no seamless process between cause and effect.

The first section of this chapter distinguishes four positions on causation based on differing beliefs about the epistemological status of mechanisms and the ontological status of the unobservables that constitute mechanisms. In the second section, I compare these four positions, and determine some methodological implications that stem from differences between these positions. The third section develops the notion of epistemic thresholds and the role they play in determining certain aspects of causation. The conclusions sum up the SSR approach to causation developed in Part II of this thesis.

i. Four Positions on Causation

The epistemological and ontological status of theoretical mechanisms each breaks down into two fundamental positions. First, epistemologically, mechanisms can be considered either *relevant* or *irrelevant to scientific inquiry*. Second, ontologically, the unobservable entities and processes that constitute mechanisms can be considered either *literally*

true/false or *fictional*. Where one stands vis-à-vis the epistemological and ontological status of theoretical mechanisms will determine whether or not one believes in causation. Table 3 captures four philosophical positions on causation that follow from this breakdown.

Humean empiricists (HEs) occupy Quadrant 1 in Table 3. HE holds that “causation reduces to regularity.” Mechanisms have no role to play in causal inquiry, because nothing intelligible can be said about the unobservable entities or processes that constitute them. In short, mechanisms are irrelevant to scientific inquiry because unobservables are *useless* fictions. HE eschews causation because empirical regularities cannot be explained by theoretical mechanisms.

Instrumentalist empiricists (IEs) occupy Quadrant 2. Unlike HE, IE contends that mechanisms have a role to play in scientific inquiry. However, the unobservable entities and processes that constitute mechanisms are only *useful* fictions that allow scientists to make predictions: unobservables have no truth value. Mechanisms are thus relevant to scientific inquiry in the sense that useful fictions can be constructed to help *predict* empirical regularities (or other observable phenomena). But IE ultimately rejects causation because, “What does not exist, cannot cause” (Miller 1987, 382).

Humean realists (HRs) occupy Quadrant 3. HR contends that unobservable entities and processes have a truth value. But it holds that even at the level of unobservables, “causation reduces to regularity.” Mechanisms are irrelevant to scientific inquiry even if

unobservable entities are literally true. HR does not hold a belief in causation because unobservable regularities are ultimately unexplained.

Finally, *scientific realists* (SRs) occupy Quadrant 4. SR holds that unobservable entities and processes have a truth value and that mechanisms are relevant to scientific inquiry. Of the four philosophical positions in Table 3, SR is the only one to hold a genuine belief in causation. Only SR holds that empirical regularities are explained by theoretical mechanisms, which, in turn, are constituted by literally true (or false) unobservables.

In what follows, I describe each of these four philosophical positions in Table 3 in greater detail. I devote a relatively small amount of space to HE, because Chapter II already outlined this approach to causation. I then move to SR because it stands in direct opposition to HE. I devote a lot of attention to SR, because this discussion constitutes the basis for the SSR approach to causation. I also devote a relatively large amount of space to IE, which is discussed next, because this approach to causation is under-appreciated in existing SR literature on causation and because it merits comparison to SR. Finally, I devote the least amount of space to HR, because I contend that it cannot cope with certain aspects of the social sciences even though it might have great relevance for the natural sciences.

Quadrant 1: Humean Empiricism

HE was explicated in Chapter II. I offer a brief reworking here simply to make this philosophical position comparable to the other three positions discussed below.

For HE, knowledge derives solely from sense experience. Consequently, unobservable entities and processes are a *useless* fiction: they have no place in natural or social science because scientists do not have special access to the unobservable world. And because unobservable entities and processes are useless fictions, mechanisms are irrelevant to scientific inquiry. The necessary connection binding observable regularities – i.e., the mechanism – cannot be discovered in any meaningful and scientific sense.

Hume's philosophy evidently epitomizes HE. For Hume (1975 [1748], 74), the unobservable entities and processes that appear to form a necessary connection between cause and effect, "are absolutely without any meaning, when employed either in philosophical reasonings or common life." Thus, HE eschews causation. To the extent that the term "causation" has any meaning whatsoever, it refers solely to empirical regularities.

Quadrant 4: Scientific Realism

SR stands in direct opposition to HE: SR holds a belief in causation because it contends that mechanisms are relevant to scientific inquiry and the unobservable entities and processes that constitute mechanisms have a truth value. But what gives SR grounds for holding these beliefs? If mechanisms are relevant to causal explanation then there must be a way to demonstrate that particular mechanisms operate as theorized. And if the unobservable entities and processes that constitute these particular mechanisms have a

truth value, there must be a way to judge whether these theoretical mechanisms are literally true or false.

My arguments in Chapters VI and VII already suggest a *partial* source of SR's belief in causation: namely, there are three methods – process-tracing, rational choice modeling, and interpretation – that can identify unique properties of theoretical mechanisms and generate findings that explain regularities in a unique fashion. And my arguments in Chapter VIII suggested a *general* source of belief in ontological status: namely, that ideas can be real and do not require an idealist ontology. But what gets SRs beyond *partial* belief that mechanisms are operating as theorized and a *general* belief that social unobservables have a truth value is the argument from coincidence (AfC).

The AfC, which is central to scientific realism (Cartwright 1979, Dessler 1991, Wendt 1999), is variously called *inference to best explanation* (IBE), *abduction*, and *the miracles argument*.¹²⁸ Although there may be some differences between these versions, the basic thrust of each is that an unobservable mechanism is tested against a range of observable phenomena, including (but not necessarily limited to) observable regularities. If the unobservable mechanism can successfully explain and predict a *wide range* of *independent* observable phenomena, SRs argue, it would be an unlikely coincidence if the hypothesized mechanism turned out to be false. The more observable phenomena the mechanism can predict and retrodict – and the greater the range and independence of these observable phenomena – the greater the coincidence would be if the unobservable

¹²⁸ Less commonly, it is sometimes also called *pattern-matching*. See Gerring (2004, 348).

mechanism were not operating as hypothesized. Confidence in a mechanism, in turn, increases our confidence that the observed regularities are genuinely causal: when we know *why* regularities occur, we are more certain that they represent genuinely causal relationships.

Dessler (1991, 349-51) contends that Charles Darwin employed the AfC to defend his theory of evolution. In effect Darwin argued that, given how well heritable variation and natural selection accounts for the evolutionary record, it would be nothing short of a large coincidence if his theory were entirely mistaken. Darwin's theory could at once explain a diverse range of phenomena, such as isomorphisms between widely different organisms, embryological similarities across species, and the geographic distribution of living organisms.¹²⁹ Thus Darwin (quoted in Dessler 1991, 350) claims, "I must freely confess the difficulties and objections [to the theory of evolution] are terrific; but I cannot believe that a false theory would explain, as it seems to me it does explain, so many classes of facts."

¹²⁹ An example of an isomorphism (species that have different ancestry but similar appearance) that Darwin's theory can explain is the common bone structure in man, bat, horse, and porpoise. An example of an embryological similarity that Darwin's theory can explain is the similarity in the embryos of mammalia, birds, and snakes. An example of how Darwin's theory can explain the geographic distribution of living organisms is its ability to explain the presence of bats, but not terrestrial mammals on oceanic islands (Dessler 1991, 350).

It is important to specify the role that regularities can play in the AfC, in order to distinguish the AfC's logic from a Humean logic of causation. Sometimes the observable facts against which an unobservable mechanism is tested are, in fact, regularities. Thus the molecular theory of matter is tested against observable regularities such as the laws of Boyle and Charles. But if the AfC relies on regularities, it is by no means a Humean method.¹³⁰

With the AfC, the case for causation does not rest on any one particular regularity, regardless of how large the N. Rather, it is the *accumulation* of evidence across a diverse range of independent regularities that provides evidence for a mechanism, and hence causation. *Hume clearly would not have suggested that we gauge causation between two billiard balls by observing, say, two ping-pong balls or two rubber balls collide. But this is precisely the type of reasoning involved in the AfC, because the very same physical forces (i.e., mechanisms) that underlie the regularity between colliding billiard balls may also underlie the regularity between colliding ping-pong or rubber balls.*

¹³⁰ Indeed, the AfC may not involve regularities whatsoever. Ironically, DSI (1994, 11-12, 223) provides a good example. It notes that if a meteorite caused the extinction of dinosaurs (as many scientists believe) then one observable implication of this theory would be the presence of iridium at certain layers in the earth's crust. The detection of iridium at precisely the predicted locations has bolstered support for the theory. The logic of causal inference employed here is the AfC – but there are no regularities involved.

With the AfC, one does not confirm the presence of a mechanism by stacking cases *vertically*: that is, by adding more instances of the *same* independent variable (IV) and dependent variable (DV) to an *existing* experiment.¹³¹ Even the most robust correlations may be spurious; adding more observations may not improve the likelihood of discovering spuriousness if the same systematic force that is creating the illusion of a causal relationship is at work with each and every additional observation. For instance, we might observe that a particular gas regularly expands when heated and appeal to the molecular theory of matter to explain this phenomenon. But we do not gain significantly more confidence in the molecular theory of matter by, say, adding *yet another* instance of this particular gas expanding when it is heated. Rather, we gain confidence in the molecular theory of matter when we move *horizontally* – that is, when the theory is tested against a set of independent phenomena in a series of *new* experiments (involving, for example, *other* types of gases, or even better, entirely independent phenomena such as that relating to Brownian motion).¹³²

¹³¹ I use the term ‘experiment’ loosely here. By experiment I include quasi-experimental methods like correlational analysis. I also include the technique of retrodiction – i.e., testing a theory against an ‘existing’ correlation, or even, an observable fact.

¹³² Brownian motion refers to a particular type of particle movement, first noticed in 1827 by the English botanist, Robert Brown, while observing pollen grains suspended in water. In the example at hand, moving horizontally to Brownian motion may be preferable to moving horizontally to a different type of gas: if *all* gases share certain properties, then they may all be subject to the same confounding factors. More generally, moving horizontally to independent phenomena is preferable to moving horizontally to phenomena that share key characteristics.

The AfC and The SSR approach to Causation

For purposes of the SSR approach, three further clarifications/refinements surrounding the logic of the AfC are required: (1) where the AfC stands vis-à-vis verification, falsification, and competing hypotheses; (2) how mechanisms identified by process-tracing fit into the AfC; (3) what constitutes a “new” social scientific experiment in the AfC.

Verification, Falsification, and Competing Hypotheses

On its surface, the logic of the AfC might appear verificationist: that is, it might seem to suggest that in order to verify a proposed theoretical mechanism, researchers should look for facts that confirm that mechanism. In fact, the argument from coincidence is not necessarily verificationist. Consistent with falsificationism, I contend that researchers should concentrate on those observable implications of their theory that are most likely to cast significant doubt on the plausibility of a proposed theoretical mechanism. However, I do not hold a *naïve falsificationist* view. A theoretical mechanism can never be confirmed, but one or two pieces of contradictory evidence may not be enough to falsify a proposed theoretical mechanism.¹³³

¹³³ Green and Shapiro (1994) contend that a “pathology” of rational choice is to search for confirming rather than falsifying evidence. Like Green and Shapiro, my approach suggests looking for the hard tests; but unlike what Green and Shapiro appear to suggest, I follow Quine (1961, 20-46) in arguing that “disconfirmed” theories are usually adjusted and reapplied.

Multiple theories, of course, can account for the same empirical phenomena. It is not enough to test a single hypothesis against empirical data; competing hypotheses must also be ruled out. As Koslowski (1996, 14) notes, “Explanations are not evaluated in isolation; they are judged in the context of rival alternative accounts.” The issue of competing hypotheses ties in with falsification: in general, the most compelling theoretical mechanism is the one with the least evidence against it, rather than the one with the most evidence for it.¹³⁴ We shall see shortly, however, that the plausibility of a mechanism in the SSR approach also depends on the extent to which it reflects the nature of real-world processes.

Process-tracing, Mechanisms, and the Argument from Coincidence

As pointed out several times throughout this thesis, there is a potential problem with my argument for mechanism that relates to the property of physicalness. If a mechanism consists of a series of regularities, it would appear that Humeans are right that causation reduces to regularity with no role for mechanism.

¹³⁴ This argument assumes that the *Quine-Duhem under-determination thesis* – the notion that for every scientific theory there will always be at least one rival theory that is equally supported by the empirical evidence – does not necessarily apply. Although I recognize the phenomenon of *equifinality* – whereby multiple mechanisms explain the same outcome – each mechanism should have at least some unique observable implications, thus making it possible to distinguish between mechanisms.

As noted in Chapter V, the Humean argument *does not* work against the second and third ontological properties of mechanisms: agency and intentionality. Neither the aggregation nor strategic interaction of rational choices, nor intersubjective meanings, are comprised of regularities. I will argue here that the Humean argument also does not apply to the property of physicalness even though process-tracing generates findings that are ultimately based on correlational evidence.

Even if a social mechanism consist of a series of regularities, the logic of the AfC treats the discrete regularities as a *single process* or mechanism from which observable implications are derived and empirically tested. Consider the meteorological process of *adiabatic lifting*¹³⁵ that Dessler (1991, 342-4) uses to illustrate his own SR approach to causation.¹³⁶ Adiabatic lifting is a theoretical mechanism that meteorologists use to explain thunderstorms. Dessler notes that adiabatic lifting explains a wide array of independent correlations, including the correlation of thunderstorms with the presence of mountain ranges, the meeting of cold and warm air masses, and hot and humid days. Dessler correctly notes that meteorology's belief in adiabatic lifting is attributable to the logic of the AfC¹³⁷: a theoretical mechanism is proposed and subsequently tested against its observable implications.

¹³⁵ Meteorologists sometimes refer to this process as *adiabatic cooling*.

¹³⁶ See Chapter III for more details on Dessler (1991).

¹³⁷ Dessler never uses the term, AfC. But the logic of testing he describes is exactly the same.

Dessler describes the mechanism of adiabatic lifting as a single process, and indeed this is accurate insofar as it describes how the mechanism is used in the AfC: meteorologists derive observable implications from the mechanism *in its entirety*. But in describing adiabatic lifting as a single process, Dessler masks the extent to which adiabatic lifting is a mechanism underpinned by correlational evidence.

Adiabatic lifting works as follows. The surface of the earth absorbs solar energy and stores this energy, causing the ground to heat up. The air over the surface of the earth is now cooler than the surface, causing a transfer of thermal energy (heat) from the surface to the air. As the air is heated, it expands, its density decreases and it becomes buoyant. At the same time that thermal energy is being transferred from surface to air, water is also evaporated from the surface to the air. Air containing a great deal of water vapour is less dense than dry air, and so the air near the surface is much less dense than the air above it. Consequently, the vaporous air rises. As it rises, the air expands and cools, gaining potential energy and losing thermal energy. The end result of this entire process is condensation and rain.¹³⁸

We can see from this description that the mechanism of adiabatic lifting actually involves a number of discrete mechanisms or causal steps. Meteorologists draw on *correlational-based physical laws* to support the discrete mechanisms that comprise the theoretical mechanism of adiabatic lifting. For example, the transfer of energy from the surface of

¹³⁸ Private communication with Tony Price, physical geographer at the University of Toronto, Wednesday, September 12th, 2002.

the earth to the air above is based on elemental physical laws, including the 2nd law of thermodynamics, and the law of conservation of energy.¹³⁹

We can conclude from this example that even if mechanisms are based on a series of correlations, this does not mean that causation reduces to correlation as the Humean might argue. Rather, the disparate correlations are *treated* as a single process or mechanism from which observable implications are derived and tested accordingly. In this sense, the concept of mechanism is pertinent.

In the social sciences, it is well recognized that process-tracing consists of linking events based on correlational evidence (usually “laws” of psychology or economics).¹⁴⁰ But sometimes, mechanisms identified by process-tracing can be tested as suggested by the SSR approach: the discrete correlations can be treated as a single process from which

¹³⁹ Dessler (1991) states that, in employing adiabatic lifting, meteorologists are not “focused” on the “presentation of empirical data supporting specific regularity claims – that is, on correlational analysis...” This is true, but misleading. The key word here is ‘focused’. Meteorologists do not *focus* on the presentation of empirical data supporting specific regularity claims because the laws they draw upon are considered beyond reproach. Few scientists would question the truth of the 2nd law of thermodynamics, and so there is simply no reason to *focus* on these regularity claims. But Dessler’s interpretation belies the fact that the theoretical mechanism is based on regularities.

¹⁴⁰ Roberts, for instance, argues that historians (he treats history as a social science) trace a path between *explanans* and *explanandum* by means of “micro covering laws”: rough and ready laws that are well established and used by lay and academics alike. See also Little (1991) and George and Bennett (2005, 224-32).

observable implications are derived and tested against the empirical record. This often consists of drawing out scope conditions from a process-tracing study. In an example discussed at greater length in Chapter XIII, Elman (1997, 191-232) uses process-tracing to analyze the decision-making process that led Finland to ally itself with Germany in World War II. Critics of the democratic peace hypothesis sometimes highlight this episode because it is an example of a democracy warring with fellow democracies. Elman's process-tracing analysis demonstrates that it was a particular brand of democracy with power centralized in the executive branch that allowed the Fins to decide to ally themselves with the Germans. Thus, Elman's process-tracing analysis suggests a possible scope conditions for further statistical tests: namely, degree of centralization in democratic institutions.

But scope conditions are not the only type of observable implication that can be drawn for process-tracing studies. A causal chain of events or phenomena, when taken in its entirety, can also suggest ways to measure (or not measure) variables (George and Bennett 2005, 48-49). As an example of the latter, Sambanis (2004, 264) argues that his case study analysis of Kenya demonstrates the problem with existing statistical models of civil war that use GDP per capita to measure state strength. Sambanis uses process-tracing to demonstrate the manner by which the Kenyan government has used corruption to exercise state control. The process-tracing analysis demonstrates that Kenya has a weak economy but a strong state, thereby calling into question the judiciousness of using GDP per capita as a measure of state strength.

There are likely other types of observable implications that can be drawn from mechanisms identified by process-tracing. But the point should be clear that in drawing these implications and testing these implications, one looks to the mechanism in its entirety. The fact that the mechanism can be parsed into discrete regularities is true, but only trivially so: by the logic of the AfC, it matters little how a mechanism is broken down because a test of that mechanism does not consist of testing each of the individual correlations that are thought to comprise the mechanism.¹⁴¹

What is a “New Experiment” in the Social Sciences?

The third and final refinement of the SSR approach to the AfC relates to the notion of moving horizontally to new experiments. This research technique needs to be specified more exactly for the social sciences. In the social sciences, one can move horizontally to new experiments by doing one or more of the following: (1) find different measures of an IV and/or DV and apply them to the same empirical domain; (2) maintain the same measures of an IV and/or DV but apply them to a new empirical domain; (3) find different measures of an IV and/or DV and apply them to a new empirical domain; and (4) find an entirely new IV and/or DV in a new empirical domain.

It must be noted that horizontal movement is “measured” by *degrees*. For example, moving to a new empirical domain constitutes a greater horizontal movement than

¹⁴¹ This is not say that the discrete steps in a causal mechanism cannot be broken down and tested by correlation analysis. (For more on this approach, see DSI.) I only say here that by the logic of the AfC, this type of correlational testing does not play a role.

finding a new measure of an IV and/or DV. This point is relevant because, as a general rule, it is preferable to move horizontally as far as possible, *because non-independent correlations are more likely to be subject to the same confounding factors.*

Consider, for instance, a hypothetical relationship between *political involvement* (IV) and *financial success* (DV). We might hypothesize a correlation based on the mechanism of personal connections: political involvement fosters personal connections, which, in turn, open up financial opportunities. We could test this by finding different measures of the IV – say, involvement in national politics and involvement in local politics – and correlating them to the DV in separate experiments. But both experiments might be subject to the same confounding factors. For example, the type of person who might become involved politically, whether at the national or local levels, might already be predisposed to financial success for other reasons. A better strategy, if possible, might be to move to a new empirical domain. For instance, looking at bureaucracies, one could test whether involvement in bureaucratic politics (such as bureaucratic associations) correlates to success (in terms of more supervisory powers, higher salary, etc.).

There are two important caveats, however, to this general rule about horizontal movement. First, in choosing alternate measures and empirical domains, one is bound by the mechanism in question. For example, the number of times a person votes might seem like a good way to measure political involvement. But measuring the number of times a person votes does not follow from the hypothesized mechanism relating to personal connections. We would *not* expect that a person who gets out to vote often would, in the

process of voting, make personal political connections. (In fact, political parties are usually barred from lobbying in voting areas on voting days because this is seen as political interference.) If there is a connection between the number of times a person votes and financial success, it is likely spurious, not causal: a third factor, namely willingness to engage in the political process might cause a person to vote and also ultimately produce financial success (through the mechanism relating to personal connections). In short, only measures and empirical domains that follow from the hypothesized mechanism should be considered.

In relation to a similar issue, George and Bennett (2005, 13, 173-5, 177-8) warn against the dangers of what Sartori (1970) calls *conceptual stretching*.¹⁴² Conceptual stretching occurs when a concept is “broadened beyond the point at which at least one relatively precise connotation (property or attribute) is retained” (Sartori 1970, 1042). Sartori’s worry is that political scientists are apt to “climb the ladder of abstraction” until a concept is so vague and fuzzy as to become, in essence, meaningless.¹⁴³ George and Bennett

¹⁴² George and Bennett actually raise this objection in relation to DSI’s techniques for increasing observable implications of a theory. These techniques are discussed at length in Chapter XI. In particular, George and Bennett claim that DSI’s second and third techniques for increasing the N of a small-N study are problematic. However, it is clear that their objection to DSI has direct implications for my arguments about the AfC: After all, by the logic of the AfC, a researcher should look – as DSI advocates – for a wide range of observable implications of their theoretical mechanisms.

¹⁴³ Sartori’s foremost concern is that statisticians will attempt to operationalize highly abstracted concepts without a firm understanding of exactly what it is they are measuring.

contend that a similar type of stretching can occur if a theoretical mechanism is applied to empirical phenomena to which it has little relevance. Looking to wide-ranging observable implications, George and Bennett argue, could “alter the research objective of the study – and, indeed, the theory itself...” (173-4). “The consequences of stretching to get all possible observable implications are not trivial,” warn George and Bennett (175).

George and Bennett’s warning about “stretching” reminds us that “the search for observable implications should be confined to those clearly relevant to the original theory” (174). Moreover, their warning points to avenues for future research aimed at developing clear guidelines for political science researchers on looking for observable implications of a theoretical mechanism.¹⁴⁴ At issue, as George and Bennett make clear, is what exactly gets to count as “relevant.”

George and Bennett’s warning about “stretching” notwithstanding, researchers need not restrict testing their theoretical mechanisms only to the main IV and DV in question. George and Bennett themselves provide an excellent illustration of moving horizontally to new empirical domains without “stretching” a theoretical mechanism out of shape. In making methodological suggestions for future research on the democratic peace hypothesis, they state:

¹⁴⁴ George and Bennett (2005, 175) note that DSI provides “[l]ittle guidance...for distinguishing between genuine, questionable, and highly speculative implications of a theory.” My account of the AfC likely does not fare much better than DSI on this account, and, as mentioned above, points to an avenue for future research.

...[I]f norms and institutions affect the international use of force, they should also affect the conditions under which domestic police forces are allowed to use deadly force.

William Hoelt [1993], for example, has argued that the domestic police forces of democratic states are more likely to be allowed to use deadly force only to prevent the use of such force against themselves or others, whereas nondemocratic state allow the use of deadly force and of state-sanctioned executions for property crimes. (58)

Moreover, George and Bennett (2005, 174) do appear to acknowledge the epistemological importance of looking for observable implications of a theoretical mechanism: “We should make clear that the reservations expressed here do not question the *general* desirability of attempting to identify observable implications *of a given theory*, both within and among cases, in order to facilitate the task of assessing *it*” (emphases in original). Indeed, given the central importance of the AfC to SR, and given the central importance of SR to George and Bennett’s methodological prescriptions, it is difficult to see how they *could* deny the importance of looking for observable implications. Their warning, however, as noted, does serve as a useful corrective to the practice of looking to observable implications that are not relevant to a given theoretical mechanism.

There is a second important caveat to the general rule, above, which states that it is preferable to move horizontally as far as possible, because, non-independent correlations are more likely to be subject to the same confounding factors. When a mechanism is *previously established*, horizontal movement can be restricted even more closely to

empirical phenomena that are related to the main regularity in question. If the mechanism is *not* previously established, horizontal movement should extend to empirical phenomena that are unrelated to the main regularity in question. To understand why this is the case, it must be clearly seen that the AfC helps to answer two questions simultaneously. First, it can help answer whether or not a theoretical mechanism actually exists and makes sense. Do molecules exist and behave according the principles of the theory of molecular motion? Second, it can help answer whether a theoretical mechanism accounts for a particular regularity. Does the theory of molecular motion account for the regularity between the heating of a gas and its expansion?

In order to *establish* the molecular theory of motion, scientists looked to a wide range of empirical phenomena – such as Boyle’s law and Brownian motion. Similarly, as discussed further in Chapter XII, Russett and Oneal (2001) look to a wide range of empirical evidence in order to test their theoretical explanation of the democratic peace, including evidence that strays quite far from the regularity between democracy and peace. For instance, Russett and Oneal test their theoretical mechanism against evidence relating to the so-called “rally ‘round the flag” phenomenon, whereby leaders bring their nations to war as a diversionary tactic.

But if a theoretical mechanism is *already established* there is no need to look beyond empirical phenomena that do not relate directly to the original regularity in question. It is not that moving horizontally beyond this realm is wrong: it is simply unnecessary and may serve to distract researchers from their main objective: namely, determining whether

the theoretical mechanism accounts for their regularity under investigation. For instance, Mesquita and Wiener (1996) invoke the mechanism of natural selection to explain a correlation between the likelihood of civil war (the DV) and the percentage of a population that is young and male (the IV). There is no need for these researchers to demonstrate that the mechanism of natural selection works. The mechanism is already known to work. Therefore, researchers can restrict horizontal movement to empirical phenomena closely related to the regularity in question. Or consider that Homer-Dixon (1999, 74-7) identifies *resource capture* as a critical mechanism linking environmental scarcity to violent conflict in developing nations: when resources become scarce, independent political factions attempt to co-opt elites and environmental resources and extract high rents from the general population for the use of these resources. Resource capture is a special kind of *rent seeking* that is well known to economists.¹⁴⁵ Thus, resource capture in this sense is a previously established theoretical mechanism, and there is no need move horizontally to empirical phenomena far beyond the realm of scarcity-conflict linkages.

Quadrant 3: Instrumentalism empiricism

Unlike SR, IE holds that unobservables are fictional: they do not have a truth value. But unlike HE, IE holds that unobservables are *useful* fictions rather than *useless* fictions. Unobservables are *useful* because they provide ‘stories’ that help predict observable, empirical regularities. These stories, or theoretical mechanisms, can be tested against the

¹⁴⁵ In economics, rent seeking takes place when an entity seeks to extract uncompensated value from others by manipulating the economic and regulatory environment.

empirical record, much as SR does with the AfC, and therefore mechanisms are relevant to scientific inquiry. But mechanisms, regardless of how much predictive power they possess, are never literally true (or false) nor do they have any independent empirical content. Thus, IE is skeptical about causation, because what does not have a truth value cannot cause.

IE espouses a particular brand of anti-realism.¹⁴⁶ The best that science can do is offer *metaphorical* accounts of unobservable phenomena (such as the metaphor of the electron circling the nucleus) that provide “empirical adequacy.” Brown (2001, 98) provides an example to illustrate the IE claim:

It may be the case that some particular astronomical theory is true (or is false), but its truth doesn't matter, since we have no hope of being able to tell if the theory is true. The heavens are completely out of reach; the meager evidence we have (or could ever hope to have) could not possibly decide which theory is true. Consequently, the aim of astronomy should not be the truth, but rather should be mere empirical adequacy. Instead of trying to

¹⁴⁶ There are many varieties of antirealism (Brown 2001, 99). BasVan Fraassen (1980) makes a particularly influential antirealist argument. Some antirealists are postmodernists or relativists rather than instrumentalists. They generally argue that the success of science is a self-fulfilling prophecy, because the experiments and conceptual schemes used to test theories presuppose their truth.

give a literally true description of the heavens, we should try to tell stories that predict where the observable points of light will be in the sky at various times.¹⁴⁷

Thus, IE holds that "...a theory should be judged by its predictive accuracy, not the realism of its assumptions" Friedman (1996, 10).¹⁴⁸

Moreover, for IE, unobservables have no independent empirical content. Theoretical entities are defined solely by the empirical phenomena that they are invoked to explain (Hempel, 1958). For instance, the ionized molecules that scatter electromagnetic radiation at frequency f – an unobservable process – is taken to be nothing more than a white streak in a cloud chamber, which is observable to the naked eye. Thus, the unobservable process has no genuine explanatory power.¹⁴⁹

¹⁴⁷ James Robert Brown, in a private communication on May 5th, 2005, points out that the basis for this type of instrumentalism can be traced back to Plato's *Timaeus*. Plato's followers took him to advocate a view called "to save the phenomena," which is closely associated with instrumentalism.

¹⁴⁸ Friedman (1996, 10) contends that instrumentalism as formulated by Milton Friedman is common to economists, who "tend to assess the validity of generalizations that have relatively weak theoretical foundations by seeing how well they account for statistical data..." Milton Friedman's brand of instrumentalism is usually associated with his *as if* arguments: that is, theoretical mechanisms operate *as if* they were true, although their literal truth is irrelevant.

¹⁴⁹ Unlike *logical positivism*, IE acknowledges that total theory-independence is not possible: theoretical entities play a role in the actual practice of science. But IE draws a distinction between discovery and confirmation and "relegate[s] talk of unobservable mechanisms to the

Quadrant 2: Humean Realism

Humean realism (HR) holds that unobservables are literally true, but that mechanisms are irrelevant to scientific inquiry because unobservable processes are ultimately comprised of nothing more than a series of (unobservable) regularities. For instance, molecular motion is said to resemble Hume's billiard balls: there is no necessary connection between interacting molecules, just as there is no necessary connection between interacting billiard balls. Most modern-day Humeans are likely HRs, because they distinguish Hume's thesis about regularities from his thesis about unobservable entities.¹⁵⁰

But if my arguments about the second and third ontological properties of mechanisms are valid, then HR may not be a tenable philosophy of causation for the *social* sciences.

Whereas physicalness may ultimately be comprised of regularities, agency and intentionality – valid only for the social sciences – are not. As argued in Chapter VI, rational choice modeling generates findings that consist of the aggregation or interaction of individual decisions; explanations of this sort are not underpinned by regularities. And as argued in Chapter VII, interpretation generates finding about intersubjective meanings

process of discovery” (Koslowski 1996, 7); or else, “rationally reconstructs” the practice of science so that when scientists appeal to unobservables they are themselves engaged in nothing more than a form of induction (Koslowski 1996, 8-9).

¹⁵⁰ This is based on several private communications with Professor James Robert Brown, University of Toronto. For an example of a philosophical approach underpinned by Humean realism, see Earman (1986, 80-110, especially p.105).

– most notably, norms and values – that provide an enabling social context.

Intersubjective meanings themselves provide explanations of regularities, but do not necessarily reduce to a series of regularities.¹⁵¹

Thus, although HR might be widely applicable to the natural sciences, I maintain that it is less applicable to the social sciences. Consequently, in what follows, I focus primarily on HE, IE, and SR.

ii. Causal Inference: The Difference that Scientific Realism Makes¹⁵²

In this section I compare SR to HE and IE, and discern some methodological implications for causal inference that stem from differences between these philosophical positions.

The comparison between SR and HE is stark and fairly straightforward. The comparison between HE and IE, however, is subtler and requires greater attention.

Scientific Realism vs. Humean Empiricism

Consider a stylized example to contrast the SR and HE logics of causal inference.

Suppose that a researcher's preliminary investigation shows a correlation between *A* and *B* and that a reasonable explanation exists for this correlation. The question arises, "Does *A* cause *B*?"

¹⁵¹ As argued in Chapters VII and IX, it may sometimes be possible to explain agency and intentionality in terms of physicalness. The point here is that in and of themselves, agency-based and intentionality-based mechanisms are not comprised of regularities.

¹⁵² "The difference that realism makes" is a phrase used by Wendt (1999).

The HE proceeds to gather more observations on both *A* and *B*, and possibly controls for more potentially confounding factors, in an effort to assess the status of the correlation.¹⁵³

If the correlation continues to hold when additional instances of *A* and *B* are stacked *vertically* to the *existing* experiment (and more controls are added), the researcher concludes that the relationship may well be causal (again, with the understanding that causation is not a genuinely meaningful term).

The SR, instead, takes more seriously the hypothesis about *why A* might cause *B*. The SR proceeds to test through correlational analysis a number of *additional* observable implications of the hypothesized mechanism – i.e., the researcher moves *horizontally* to *new* independent experiments that are entailed by the hypothesized mechanism. If the hypothesized mechanism correctly predicts (and/or retrodicts) these correlations, the researcher assumes that the mechanism is indeed operating as theorized. Consequently, they conclude with much greater confidence than was previously the case that the original correlation between *A* and *B* is causal. (Each experiment involves all the normal rules of causal inference, such as avoiding selection bias and controlling for potentially confounding factors.)

In short, the methodological implications that stem from the difference between SR and HE approaches to making causal inference are stark and relatively straightforward.

¹⁵³ The HE might also check for temporality.

Scientific Realism vs. Instrumentalism Empiricism

Like SRs, IEs use theoretical claims about unobservables to predict and retrodict empirical regularities. And as with the AfC, it is preferable to identify theoretical mechanisms that can be evaluated in a range of empirical domains. This similarity between SR and IE, however, belies a crucial difference between these two philosophical positions.

For IEs, mechanisms are *only* useful insofar as they generate hypotheses about empirical regularities (and, more generally, observable phenomena). The analysis flows one way only, from mechanisms to regularities. Unobservables are merely deductive tools or *heuristic devices* (MacDonald 2003, 553). For SRs, the analysis flows back to mechanisms: as dictated by the AfC, every additional regularity that a mechanism successfully explains increases one's confidence that the mechanism is operating as theorized and is literally true.¹⁵⁴

¹⁵⁴ Moreover, SRs point out that unobservables cannot be causal if they are defined solely in terms of empirical phenomena. Greenwood (1994, 6-7) contends that Hempel's approach does not allow for genuinely causal explanation: "If 'intervening variables' really are defined in terms of the empirical laws to which they bring conceptual integration, any putative explanation of empirical laws in terms of intervening variables would be viciously circular, since it would not embody any information not already contained in the statement of the empirical laws. Thus if 'schizophrenia,' for example, is defined in terms of a behavioral syndrome, it can hardly be advanced as an explanation of that syndrome."

But in terms of *methodology*, there might not appear to be much at stake in the debate between SR and IE. The choice between the two might seem to be based strictly on philosophical differences about what we can learn and know by testing theoretical mechanisms against a range of observable implications. This philosophical issue may be important, but what *pragmatic* difference could it possibly make?

In fact, the SSR approach has crucial methodological implications for the social sciences. The choice between SR and IE determines whether social scientists should aim for theoretical mechanisms that emphasize *generalizability* or that emphasize *accuracy*. This choice, in turn, determines how much value social scientists attribute to case study research (this applies to mechanisms identified by process-tracing and interpretation) and how social scientists should approach theoretical assumptions in rational choice models (this applies to mechanisms identified by rational choice modeling).

Generalizability vs. Accuracy

IEs favour *parsimonious* theories: the best theories are simple, elegant theories. Because IE treats unobservables as fictions, it does not matter whether or not theoretical mechanisms accurately reflect reality. All that matters, is empirical adequacy. Simple mechanisms are preferred to complex mechanisms, because simple mechanisms are portable to a wider variety of empirical domains (MacDonald 2003, 556).

SRs also want mechanisms with wide generalizability. The logic of the AfC relies on the ability of mechanisms to explain a range of empirical regularities and phenomena.

Without generalizability to an array of empirical domains, there would be no *coincidence* of which to speak. But SR also emphasizes *accuracy*. If mechanisms are to reflect reality, they must be as complex *as* that reality. SR “places a premium on the design and construction of theories that attempt an accurate description of the processes that underlie the universe” (MacDonald 2003, 554). Parsimony and generalizability are moderated by the demand for accuracy and detail. SR will sacrifice parsimony and generalizability on the altar of accuracy, because if unobservables do not reflect reality, there can be no causation (George and Bennett 2005, 31, MacDonald 2003, 556). As George and Bennett (2005, 142) argue: “At the frontiers of research...social scientists need to discard stylized simplifying assumptions and build upon the most accurate microlevel mechanisms that can be discerned.”

This does *not necessarily* mean that theoretical mechanisms *must* be complex or that the social world is inherently complex. It only means that theoretical mechanisms must be as complex as reality.¹⁵⁵ As George and Bennett (2005, 147) assert: “[C]omplexity is not intrinsic to the definition of causal mechanisms, even though mechanisms operate in historical contexts that are often complex...” The crucial difference here, between SR and IE, is that according to the SSR approach, SR eschews the simplifying assumptions and scaled-down models that typify IE.

¹⁵⁵ By the same token, favouring parsimony does not entail believing that the social world is inherently simple (King et al. 1994, 20, Gerring 2005, 174).

Another way to conceive of this divide between SR and IE is in terms of the trade-off between prediction and explanation. IE emphasizes prediction to the detriment of explanation, whereas SR emphasizes explanation to the detriment of prediction. SR holds that as explanations become more complex in order to reflect reality, predictions might actually become more accurate, but predictive *range* will be lost: again, the more complex the theoretical mechanism, the less it likely to explain.

The tradeoff between the complexity and generalizability of theoretical mechanisms raises important issues for how the SSR approach advocates choosing between competing theories. Traditionally, philosophers of science use parsimony as one (of several) criterion by which to judge between competing theories: theories that explain more with less are preferred to theories that explain less with more.

The SSR approach does not take parsimony to be a criterion by which to judge competing theoretical mechanisms. Instead, one of the most important criteria is achieving an appropriate *balance* between theoretical detail and generalizability. We do not want theoretical mechanisms that include *nonsystematic factors* – i.e., random variation and “noise.” But we also do not want theoretical mechanisms that exclude relevant *systematic factors* – i.e., factors that have some applicability across time and space. Thus, the SSR approach does not eschew generalizability – to do so would render useless the AfC. But the SSR approach recognizes that generalizability will be curtailed when theoretical mechanisms reflect the complexity of the social world. Theories that strike an appropriate balance between theoretical detail and generalizability will be preferred to

theories that are skewed in either direction.¹⁵⁶ In addition, other more traditional criteria for choosing between rival theoretical mechanisms are also applicable to the SSR approach, such as coherence with other established theories and capacity to generate novel predictions (Psillios 1999, 171).

The Value of Case Studies

The divide between SR and IE over the issues of parsimony, generalizability, and prediction has implications for the role of case study and ethnographic research in the social sciences, in particular, the methods of process-tracing and interpretation. For SR, case studies are critical tools for the development of theoretical mechanisms, because case studies can generate findings about mechanisms that reflect reality: after all, empirical research of real-world phenomena is the *sine qua non* of case-study research.¹⁵⁷ For IE, case studies might be useful insofar as they generate testable hypotheses; but case studies are not required to generate testable hypotheses. Moreover, the complexity of the mechanisms derived from case studies can be harmful to the goal of parsimony. In general, case-study research is an integral component of causal inference in the SSR approach, which is why I emphasize the importance of process-tracing and interpretation.

¹⁵⁶ The question of how one might know when the appropriate balance is struck is addressed below in this chapter.

¹⁵⁷ Critics might respond that case-study research does not reflect reality, because case-study methods –in particular, process-tracing and interpretation – are methodologically flawed. But I do not suggest that case studies necessarily “get it right.” As I have argued, further testing of theoretical mechanisms is almost always required. What I *do* suggest is that if we are going to *attempt* to capture reality, case studies are an invaluable tool.

For IE, on the other hand, case-study research is at best moderately useful and at worst detrimental.¹⁵⁸

The value that the SSR approach attaches to case studies extends beyond *developing* theoretical mechanisms. The SSR approach also holds that, to a limited extent, case studies are useful in and of themselves for *testing* mechanisms and their corresponding regularities.¹⁵⁹ By my view, the line that IE draws between the building and testing of theoretical mechanisms is too sharp: building theoretical mechanism is *itself* a form of testing. Because case-study researchers work with empirical evidence, they cannot simply conjure up any story that fits their purposes.¹⁶⁰ Although some might contend that case studies lack rigorous standards of evaluation (at least in comparison to quantitative

¹⁵⁸ I do not suggest that interesting and realistic theories *necessarily* require case study analysis. It is possible to develop abstract theoretical mechanisms that reflect reality. Rational choice theorists do this in the social sciences. But as we will see shortly, the difference between SR and IE over generalizability and parsimony also has implications for the construction of abstract theoretical mechanisms.

¹⁵⁹ For a vigorous defence of the ability of case studies to test theories, see George and Bennett (2005) and Dion (2003).

¹⁶⁰ If case study researchers work with empirical evidence, it might reasonably be asked why I include this type of research in a discussion of purportedly *unobservable* mechanisms. The answer is that case study researchers work both with the observable and the unobservable. In essence they use observable entities to make inferences about unobservable processes. Moreover, because the SSR approach treats the linkages identified by process-tracing as a single process, this single process is for all intents and purposes, unobservable.

analysis), the empirical evidence presented by case-study researchers can reasonably be judged as more or less compelling.¹⁶¹ Mechanisms cannot be developed “willy-nilly”: the relevant community of scientists will not tolerate models that are not founded on sound scientific principles. It not just that adiabatic lifting successfully retrodicts certain regularities that makes it a successful mechanism: the mechanism is also compelling because it is founded on sound scientific laws such as the laws of thermodynamics. Thus, building theoretical mechanisms is a form of scientific testing.¹⁶²

If a case study researcher is able to explain an empirical regularity by developing a compelling account of a mechanism, then we should be more confident that the regularity

¹⁶¹ Following Friedman (1996), I contend that theoretical mechanisms can and should be reformulated after being disconfirmed through empirical testing. However, the process of reformulating a mechanism should be transparent; and once reformulated, it must be acknowledged that a new mechanism is ready for further testing. Failing to take these steps can lead to what Green and Shapiro (1994) call “arbitrary domain restriction.”

¹⁶² By the same token, scientific testing is a form of theory building. Consider that Thomas Edison essentially developed his theories by trial-and-error experimentation. More generally, failed experiments send researchers “back to the drawing board” where they tweak their theories. Koslowski (1996, 10) points out that “...just as theory can enhance method, so method can lead to new discoveries that can refine and amplify theory. Or, as Gerring (2004, 350) puts it: “Traditionally, scientific methodology has been identified with the segregation of conjecture and refutation... Yet in the real world of social science, inspiration arises from perspiration.” Babbie (1998, 59) refers to this cycling between induction and deduction as “the wheel of science.” Diesing (1991) makes a similar argument.

is causal. (But, of course, even compelling mechanisms must be subject to further empirical tests.) Conversely, failure to develop a compelling explanatory mechanism casts doubt on the causal nature of a regularity. As noted in Chapter III, the bacterial theory of ulcers was not taken seriously when first proposed precisely because there was no compelling mechanism. Lipton (1992), recounts that within a few hours of anthropological field work in a single village, he managed to raise serious issues about widely regarded statistical-based relationships. Again, a single case can rarely in itself overturn a strong correlation, but it may raise questions about the causal nature of a correlation.¹⁶³ This would explain why, in actual scientific practice, theoretical mechanisms not borne out by empirical evidence are not necessarily rejected outright. Indeed, if the theoretical mechanism is compelling and based on sound scientific principles, the empirical tests themselves may be reformulated: “[S]ometimes theoretical considerations dictate rejecting (or at least questioning) the data themselves...especially when the hypothesis in question is especially theoretically plausible...” (Koslowski 1996, 10-1).¹⁶⁴

¹⁶³ Some political scientists contend that a single case can overturn a theory (e.g., see Eckstein 1975). Rogowski (1995, 467-8) claims that Arend Lijphart’s 1968 study of the Netherlands refuted David Truman’s theory of “crosscutting cleavages.” I would argue that single cases rarely, if ever, overturn theories, a position that follows from Quine (1961, 20-46). Gerring (2004, 350, 2005, 184) makes a similar argument about single cases in political science research. Lijphart (1971, 691-2) himself argues that single cases cannot entirely overturn a theory, but can be used to re-develop an existing theory.

¹⁶⁴ Case studies can also be used to test existing theories. As George and Bennett (2005, 217) note about process-tracing: “If a theory is sufficiently developed that it generates or implies

Rational Choice Modeling

I have argued that the divide between SR and IE over generalizability and parsimony has implications for the relevance of case-study research, in particular, process-tracing (which relates to the first ontological property of social mechanisms, physicalness) and interpretation (which relates to the third ontological property of social mechanisms, intentionality). Whereas IE holds that case-study research is inconsequential (and perhaps even detrimental), SR holds that case-study research is critical to both the development and testing of mechanisms. The differences between SR and IE have somewhat different consequences for rational choice modeling (which relates to the second ontological property of social mechanisms, agency): at issue is not whether rational choice modeling should be used by social scientists, but rather how rational choice models should be constructed. MacDonald (2003, 556) makes this argument:

Because of their emphasis on simple, elegant, and testable hypotheses, [rational choice] theorists who subscribe to instrumentalist-empiricism¹⁶⁵ should push for theoretical assumptions that increase the generalizability of their hypotheses. Conversely, because of their focus on causal

predictions about causal processes that lead to outcomes, then process-tracing can assess the predictions of the theory. In this use, process-tracing evidence tests whether the observed processes among variables in a case match those predicted or implied by the theory.” George and Bennett do qualify this assertion in several important regards (see especially, p.222-3).

¹⁶⁵ I have adopted this term from MacDonald, but altered it just slightly to read either “instrumentalism empiricism,” or “instrumentalist empiricist.”

mechanisms, [rational choice] theorists who subscribe to scientific-realism should favor theories whose assumptions create realistic hypotheses, even at the expense of parsimony or testability.¹⁶⁶

MacDonald (2003, 556-9) argues that rational choice theorists must choose between SR and IE because this choice has implications for how to treat some core theoretical assumptions underlying rational choice modeling: namely, rationality, self-interest, and methodological individualism.

Rationality

Do individuals always behave rationally? Do they, for instance, consistently maximize expected utility? Do they always behave purposively? Questions like these about the universal applicability of the “rationality assumption” divide rational choice theorists and their critics. MacDonald contends that how one responds to these questions, depends on one’s philosophical approach.

A SR approach to rational choice modeling espouses the *domain response*, whereby the assumption of rationality is restricted to particular circumstances or contexts. Although the domain response naturally limits the universality of rational choice modeling, it more accurately reflects reality – which is key because “[f]or scientific realists, the rationality assumption is intended to capture the actual cognitive processes that are involved in

¹⁶⁶ I agree with MacDonald on the issue of *parsimony*, but *not* on the issue of *testability*. Contra MacDonald, I argue that SR like IE, *always* strives for testable hypotheses: the AfC requires testable hypotheses.

human decision-making” (MacDonald 2003, 557). An IE approach to rational choice modeling, conversely, espouses the *as if response*, whereby individuals “behave as if they were following the dictates of the rationality assumption, even if actual decision making proves otherwise” (556). The “as if” response preserves the universality of rational choice modeling. It matters little to IE if individuals’ decisions are sometimes motivated by factors other than rationality. All that matters to IE is empirical adequacy. As long as, on average, the assumption of rationality aids in prediction and retrodiction, the reality of the assumption is irrelevant.

Self-Interest

Similarly, the choice between SR and IE has implications for the assumption of self-interest. Rational choice theorists are divided over whether individuals are sometimes motivated by factors other than self-interest – altruism or religious belief, for instance – or whether self-interest is universally applicable.

SR generally espouse thin rationality, whereby individuals “...possess preferences over almost anything – including things external to the agent, such as material goods, and things internal to the agent, such as emotional satisfaction and other nonmaterial values” (557). IE generally espouses thick rationality, whereby individuals are modeled “with the assumption that they all maximize the same set of consistent preferences...,” and this set reflects narrow self-interest (557). Again, SR limits the applicability of the rational choice modeling but is more accurate, whereas IE assumes universality regardless of

whether the actual preferences that rational choice modeling is assumed to represent are literally true.¹⁶⁷

Methodological Individualism

Finally, the distinction between SR and IE has implications for how rational choice theorists treat *methodological individualism*. Rational choice theorists generally agree that "...macrosocial outcomes are the sum of discrete, intentional acts..." but disagree whether "[v]ariables that cannot be reduced to the individual actors – such as the arrangement of the actors in relation to one another or the environment of their interaction..." should be included in rational choice models (558).

SR holds that rational choice modeling "can incorporate structural position and the constitution of actors into its purview" (558). This approach limits the applicability of rational choice modeling but more accurately reflects actual cognitive processes and interactions of individuals. IE generally rejects the notion that "rules of the game" can be "introduced exogenously" into its models, arguing instead that structural forces can be largely ignored and the "rules of the game" stem solely from the "interactions of individuals" (559). This approach "emphasizes generalizability, but does so at the

¹⁶⁷ Defending a realist approach to rational choice modeling, Friedman (1996, 4) argues, "The extent of self-interestedness is therefore likely to vary historically as perceptions of appropriate behaviour change."

expense of the realism of the theory, given that many effects observed in social life are likely to be the result of structural factors” (559).¹⁶⁸

Reality and Mechanisms

A key point in the above arguments is that SR promotes the development of mechanisms that reflect reality. This is why SRs emphasize the usefulness of process-tracing and interpretation, and “realistic” modeling of rational decisions. Theoretical mechanisms that are proposed merely because they seem to make intuitive sense are a good beginning but are usually sparse and underdeveloped: i.e., they do not reflect reality. As will be seen in Part III, theoretical mechanisms constructed to explain the *democratic peace* are often of this nature. A SR approach to causal inference involves deepening our understanding of the explanatory processes by examining them empirically with case studies or through “realistic” rational choice modeling.

iii. Causation and Epistemic Thresholds

I have argued in this chapter that unlike HE, HR, and IE, SR holds a genuine belief in causation; and this belief has methodological implications for the social sciences. But this chapter and the previous chapter raise two related problems for the SSR approach that must be addressed. In the previous chapter I argued that mechanism are always in need of further explanation. This raises two related questions: “How can a researcher

¹⁶⁸ See also Forbes (2004, 62) on the “rules of the game” and the inevitable tradeoffs that an approach based in realism yields.

know when a regularity has been sufficiently explained?"; "At what point can the regularity be said to be causal?"

Similarly, this chapter has argued that a crucial feature of SR – something that helps distinguish it from IE – is that, although it emphasizes generalization, it does not do so at the expense of accuracy in the description of theoretical mechanisms. This point also raises three additional questions: "How can a researcher know when he or she has struck the correct balance between theoretical accuracy and empirical generalizability?"; "Is there a particular balance at which point causal inference can be made?"; "In choosing between competing theoretical mechanisms, how can we know which one has struck the more appropriate balance between theoretical detail and generalizability?"

The answer to both sets of questions is unlikely to satisfy those seeking rigid rules for causal inference. There are no equivalencies to, say, rules pertaining to statistical significance. No formal methods or techniques exist to answer these questions.

Developing these rules may be an avenue for further research, but my provisional response is that epistemic communities in their collective wisdom determine the answers to both sets of questions. Collectively, communities of researchers determine when mechanisms are no longer in need of further explanation and when the appropriate balance between theoretical detail and generalizability has been struck.

The SSR approach therefore holds that causation is gauged, in part, by epistemic confidence. There is what I will call a *confidence level in causation*, or a "CLIC,"

beyond which deeper explanation is not required to establish causation, and the appropriate balance between accuracy and generalizability has been established. A scientific community converges around a particular empirical regularity and, in Popperian fashion, moves closer and closer to causation as that regularity is explained at successively deeper levels of explanation. Causation, like objective truth, lies behind us at some unknown point. The best we can do is blindly step backwards without ever knowing whether or not we have reached our destination.

A key indicator that an epistemic threshold, or CLIC, has been surpassed is when criticism from skeptics about a given theoretical mechanism subsides significantly. Thus, we have an understanding of the physiological process that causes death when a bullet strikes the heart. Deeper explanation of this process is not required to demonstrate causation because few medical scientists would doubt the physical mechanisms that explain the correlation between ‘bullets to the heart’ and death. Conversely, we will see in Chapter XII that skeptics continue to criticize theoretical mechanisms invoked to explain the democratic peace, in part because these mechanisms are perceived as overly simplistic, even if they can explain a wide range of empirical phenomena. In short, critics of the democratic peace hypothesis sometimes argue that the appropriate balance between theoretical detail and accuracy has not been struck, and hence the causal nature of the relationship between democracy and peace is called into question. In brief, whereas a CLIC has been surpassed in relation to mechanisms that explain the correlation between ‘bullets to the heart’ and death, this epistemic threshold has not been surpassed in relation to mechanism that explain the correlation between democracy and peace.

iv. Conclusions: A Summary of the SSR approach to Causation

Part II of this thesis has developed the SSR approach to causation. I have distinguished three ontological properties of social mechanisms, each of which is identified by a particular methodology. Process-tracing identifies physicalness, the first ontological property of social mechanisms, and generates findings that explain regularities as a chain of events or phenomena. Rational choice modeling identifies agency, the second ontological property of social mechanisms, and generates findings that explain regularities as the aggregation or strategic interaction of individual choices. Interpretation identifies intentionality, the third ontological property of social mechanisms, and generates findings that explain regularities as being enabled by intersubjective meanings.

Process-tracing, rational choice modeling, and interpretation can be used as partial tests to determine whether correlations are causal or spurious. Moreover, there are sound philosophical arguments for why, generally, mechanisms identified by these methods have a truth value, despite the fact that they are constituted by mind-dependent ideas. But more robust testing is required to determine whether mechanisms operate as theorized and whether specific mechanisms are literally true or false. I have argued that the argument from coincidence (AfC) provides this methodological approach. Observable implications are identified by process-tracing, rational choice modeling, and interpretation, and these observable implications are tested in an array of different empirical domains.

Generally speaking, the more empirical phenomena that a hypothesized mechanism can account for, the greater is our confidence that the set of regularities it helps explain are genuinely causal. This requirement for generalizability, however, is moderated by a requirement for theoretical detail and accuracy. Predictive range is sacrificed in order to identify theoretical mechanisms that reflect real social processes, because what does not exist cannot cause. At some point, epistemic thresholds are surpassed beyond which further testing is not required: a scientific community come to general agreement that a given mechanism is not in need of further explanation, and that a given mechanisms has struck a more appropriate balance between generalizability and theoretical detail than competing mechanisms.¹⁶⁹

Of the four philosophical positions indicated in Table 3, only SR holds that theoretical mechanisms are relevant to scientific inquiry and unobservables have a truth value, and hence only SR holds a genuine belief in causation. I contend that the SSR approach to causation, as outlined immediately above, provides political scientists with the philosophical and methodological tools required to make genuine causal inference. In Part III of this thesis, I examine how causal inference *is* made in political science, and suggest that, contra some existing accounts by SRs in the social sciences, *relatively* minor adjustments are required to fully adopt the SSR approach.

¹⁶⁹ Again, other more traditional criteria, such as explanatory coherence, should also be used in choosing between competing mechanisms.

Part III: Causal Inference In Political Science

The SRs discussed in Part I typically assert that HE characterizes political science and other social science disciplines. I argue in Part III that, in fact, an uneasy mix of HE, IE, and SR characterizes political science in both its methodological prescriptions and its actual research. I contend that by more fully adopting the SSR approach to causal inference outlined in Part II, the discipline of political science can advance on several fronts.

In Chapter XI, I demonstrate that King, Keohane, and Verba's methodological prescriptions for making causal inference in DSI are philosophically inconsistent: DSI argues that causation is a genuinely meaningful term, but its formal definition of causation champions HE, and its pragmatic advice for increasing the N of a small-N study is ambiguous: it implicitly embraces aspects of both IE and SR. In Chapter XII, I show that democratic peace research almost never bases its causal inferences in HE, but rather typically sits uneasily on the cusp between IE and SR. In Chapter XIII, I argue that by adopting the SSR approach to causal inference, the discipline of political science might: generate better-trained practitioners that can "make sense" of DSI and even extend some of its useful advice; bridge deep methodological cleavages in the discipline by acknowledging the inherent compatibility of various methodologies; and yield more precise information for policymakers by emphasizing detailed and accurate mechanisms.

Chapter XI: DSI and Causal Inference

Chapter X outlined four competing approaches to causal inference, three of which have practical, methodological implications for the social sciences. HE disavows theoretical mechanisms and focus strictly on making causal inference by correlational analysis. IE uses theoretical mechanisms to generate hypotheses about correlations, but denies that theoretical mechanisms play an integral role in making causal inference because mechanisms have no genuine explanatory function. IE emphasizes parsimonious theories and downplays the importance of developing mechanisms that reflects reality. IE prefers rational choice models that reflect this emphasis on parsimony and downplays the role of case studies. Like IE, SR uses theoretical mechanisms to generate hypotheses about correlations. But SR contends that the argument for coincidence (AfC) provides a robust methodological tool to gauge whether mechanisms operate as theorized and whether or not they are literally true. Theoretical mechanisms are genuinely explanatory and therefore integral to causation. SR emphasizes accuracy over generalizability; it places a high value on case study research and champions rational choice models that reflect this emphasis on accuracy.

Which of these three competing philosophical approaches to causal inference best characterizes mainstream political science? This is the central question of this chapter and the next. Typically, proponents of SR in the social sciences cast their opponents as HEs (e.g., see Dessler 1991 and Little 1991). I agree that SR is not in full bloom in political science, but this characterization is nevertheless only marginally accurate; greater nuance is required. In this chapter we will see that, in its methodological

prescriptions, political science exhibits HE, IE, and even SR. In the following chapter we will see that, in its actual research, political science almost never exhibits HE; instead political science research sits on the cusp of IE and SR.

In this chapter, I draw on King, Keohane, and Verba's (1994) *Designing Social Inquiry* (DSI) to demonstrate that methodological prescriptions in political science are a schizophrenic *mélange* of HE, IE, and SR. DSI is widely touted as the leading methodological and research design textbook in mainstream North American political science. In fact, DSI is nothing short of a methodological manifesto for mainstream political science: it is preached with evangelical verve in political science departments across North America.¹⁷⁰ It therefore serves as a fair work on which to judge the philosophical underpinnings of "state of the art" methodological prescriptions in mainstream political science.

The first section of this chapter examines DSI's formal definition of causation, arguing that it is consistent with HE. The second section contends that, contra its formal definition of causation, DSI implicitly emphasizes the importance of theoretical mechanism in their pragmatic advice on how to increase the N of a small-N study. The third section argues that DSI's implicit emphasis on mechanism is ambiguous: it appears to be founded on elements of both IE and SR. The conclusions suggest that DSI embodies a tension in political science: its explicitly stated desire to make *causal* inference is not complemented by the philosophical underpinnings on which to do so.

¹⁷⁰ DSI is influential in Western Europe as well, although to a lesser extent.

i. DSI and Humean Empiricism

In political science, HE emerged with the behavioral revolution, which heavily influenced the discipline of psychology from the early- to mid-1900s.¹⁷¹ *Methodological behaviorism* and *radical behaviorism* were particularly influential doctrines in psychology, and, by extension, political science.

The epistemological and ontological status of unobservables is one of the primary axes around which the history of psychology has revolved.¹⁷² Both methodological and radical behaviorism explicitly rejects the search for unobservable entities and mechanisms: science for these positions is a search for observable regularities and should not concern itself with the development and testing of theoretical mechanisms (Poling et al. 1995, Thyer 1999, Uttal 2000).

Methodological behaviorism arose in the early 1900s, in large part as a response to *introspectionism*, which proceeded on the assumption that the internal workings of the

¹⁷¹ Behaviourism, like rational choice and interpretation, can be considered a theoretically oriented method.

¹⁷² The pendulum has swung back and forth in psychology: introspectionism embraced the unobservable realm, behaviorism rejected it, and modern-day cognitive psychology has reintroduced it (although with scientific methods lacking in the introspectionist approach).

mind – mental processes, consciousness – are accessible.¹⁷³ Behaviorism’s principal founder, J.B. Watson, attempted to expunge this unobservable realm – what Watson referred to as the “ghost in the machine” – from the study of psychology (Watson 1966 [1924], 1-19). Watson (1966 [1924], 2), argued that “belief in the existence of consciousness goes back to the ancient days of superstition and magic.” He argued that psychology “can find no objective evidence for [the] existence” of theoretical entities like mind and consciousness (Watson 1966 [1924], 18). Mackenzie (1977, 26) notes that behaviorism was characterized by the “refusal to give any consideration to any entities or processes which were not directly and publicly observable, a refusal that was explicitly implemented as a methodological maxim.” As Watson (1966 [1924], 6) put it, “...Why don’t we make what we can *observe* the real field of psychology?”

Radical behaviorism, which emerged in the mid-1950s differed from methodological behaviorism in many respects but nonetheless mostly retained its rejection of a place for unobservable entities and processes in science. B.F. Skinner, the principal founder of radical behaviorism, once opined that all references to unobservable entities and processes were “exhausting digressions” (Skinner as quoted in Greenwood 1994, 86).¹⁷⁴

¹⁷³ Behaviorism was also in part a response to rapid progress in medicine, chemistry, and physics – progress that John B. Watson and other behaviorists wanted to emulate. See Watson (1966[1924], 5-6).

¹⁷⁴ Prior to 1950, Skinner was far less adverse to unobservables. Skinner, in fact, may have been the first in psychology to introduce unobservable *intervening variables*, although they were defined operationally and were not considered causally efficacious. Post 1950, however, Skinner became “ultrapositivistic” and “increasingly antitheoretical” (Hull, 1966, ix-xvii). (“Anti-

Like methodological behaviorism before it, radical behaviorism distinguished itself from introspectionism and defined its philosophy of science by its empiricist approach to unobservables.

In psychology, behaviorism has been largely eclipsed by cognitivism (in which the unobservable realm is a central feature), but it remains influential on the fringe. Books extolling and encouraging behaviorism's methodological and philosophical virtues continue to be published (e.g., Poling et al. 1995, Uttal 2000), books detailing the history of behaviorism continue to be written (e.g., Mills 1998, Thyer 1999), and journals dedicated to publishing behaviorist research continue to operate (e.g., *The Journal of Applied Behavioral Science*, and the *Journal of Applied Behavior Analysis*).

In political science, few scholars would openly embrace behaviorism. But as Dahl (1961) notes, behaviorism did not disappear from the discipline of political science because it failed but rather because it succeeded. Behaviorism emerged as a protest movement against “unscientific” practices in political science but its principles were soon absorbed by the discipline's mainstream. In this sense, behaviorism was “the first victim of its own triumph” (Dahl 1961, 119).¹⁷⁵

theoretical” here means that Skinner would not countenance unobservable entities and processes, not that he was opposed to theory.) Even post 1950, Skinner (1974, 16) claimed that radical behaviorism accepts the unobservable world but “...does not call these events unobservable...”

¹⁷⁵ See also Almond (2004) and Monroe (2004) on the lasting impact of the “Chicago school of political science,” which spearheaded the behavioural revolution in political science.

DSI, at least in part, reflects the lasting influence of the behavioral revolution. DSI's formal definition of causation is decidedly based in HE, as is at least one of their fundamental rules for making causal inference.

DSI's Formal Definition of Causation

Although DSI does not reference the philosophy of David Hume, it formally defines causation strictly in Humean terms. Causation is said to exist when, over numerous replications, a change in the value of a particular independent variable (IV) makes a difference, on average, to the value of a particular dependent variable (DV). DSI defines this difference as a *causal effect*. In short, causation is synonymous with causal effect, which is simply its term for a regularity or correlation.

I abbreviate the example that DSI uses to illustrate its formal definition of causation. DSI asks us to imagine a hypothetical election in the Fourth Congressional District in New York, with a Democratic incumbent and one Republican (nonincumbent) challenger. The goal is to assess whether or not incumbency is related to the proportion of votes a candidate receives. DSI asserts that under ideal (albeit unachievable) circumstances, we would repeatedly go back in time to the start of the election campaign – with every historical detail up until that point held constant – and run the election, first with an incumbent Democratic candidate and then with a non-incumbent Democratic candidate. The causal effect of incumbency in the Fourth Congressional District in New York is the difference between two vote fractions: the average proportion of the vote received over

numerous iterations of the election with a Democratic incumbent, and the average proportion of the vote received over numerous iterations of the election with a Democratic non-incumbent (76-82).

Notice that in this formulation there is no question as to why the status of incumbency might alter voting results. There is only a question as to whether, on average, a change in the status of incumbency is regularly associated with a change in vote totals. This is why DSI argues that its definition of causation “is logically prior to the identification of causal mechanisms” (86). In brief, causation according to DSI’s formal definition is nothing more than an unexplained regularity, which is entirely in keeping with HE and the behaviourist tradition.

ii. The Importance of Theoretical Mechanisms in DSI

It would be inaccurate to suggest that the legacy of behaviorism was to completely deny a role for theoretical mechanisms. In fact, the behaviorism of Clark L. Hull emphasized the importance of theoretical mechanisms in manner consistent with IE.¹⁷⁶ Contra the HE espoused by methodological and radical behaviorists, Hull’s behaviorism held that theoretical mechanisms could be invoked to explain empirical phenomena. As Hull (1952, 1) points out in his opening remark of *A Behavior System*, “Science has two essential aspects – the empirical and the explanatory.” For Hull (1966 [1943], xiii), theoretical mechanisms consist “of a chain of intervening variables interpolated between

¹⁷⁶ To a certain extent, other psychologists such as Skinner (pre 1950) and E.C. Tolman (1932) anticipated Hull’s position.

the independent environmental variables and the dependent response variables.”¹⁷⁷

Moreover, in keeping with IE, Hull’s theoretical mechanisms could be deduced from his general theoretical system and tested against empirical reality (Hull 1952, 351). Like all behaviorists, however, Hull denied that theoretical mechanisms have a truth value, and hence genuine explanatory power.

The extent to which Hull’s brand of behaviorism, influenced the discipline of political science is uncertain. But in DSI we do find a similar emphasis on theoretical mechanisms that coexists very uneasily with its embrace of HE. In particular, some of DSI’s advice on how to increase the N of a small-N study depends on the development of theoretical mechanisms. (The question of whether this development of theoretical mechanisms is based in IE or SR will be discussed below. The immediate point is that this emphasis on theoretical mechanisms is inconsistent with HE.)

A significant portion of DSI describes the dangers of small-N research – including, selection bias, measurement error, and endogeneity – and provides constructive techniques to overcome these obstacles. But throughout the book, readers are continuously reminded that the *best* method to overcome the problems of a small N is to increase the number of observations (the N). To this end, DSI suggests three techniques for “making many observations from few” (217-28).

¹⁷⁷ This quote is actually from Hull’s disciple, Herbert Spence, who wrote the preface to Hull’s book.

Below, I contrast the first techniques suggested by DSI for adding observations to a small-N study – “same measures, new units” –with the second and third techniques they suggest – “same units, new measures” and “new measures, new units.” I show that the first technique is, in fact, perfectly consistent with its formal definition of causation. I argue, however, that the second and third techniques are *not* consistent with HE because of the emphasis on theoretical mechanisms.

First Technique: Same Measures, New Units

The first technique suggested by DSI (219-23) is philosophically consistent with the HE of their formal definition of causation. Here, the researcher adds observations across time and space, keeping the same measures (i.e., maintaining the same operationalizations of the IV and DV) but adding new units (i.e., adding more instances of the original IV and DV). Thus, DSI (219) suggests adding “Pakistan, Bangladesh, and Sri Lanka to one’s data base along with India.”

When new units are added across time and space while maintaining the same IV and DV, causation reduces to a regularity. With this technique, observations are added by stacking instances of the IV and DV *vertically* to an *existing* experiment. In short, DSI’s first technique for “making many observations from few” is philosophically consistent with its formal definition of causation.

Second Technique: Same Units, New Measures

But contrast this first technique with DSI's second technique for adding observations to a small-N study: "Additional instances for the test of a theory or hypothesis can be generated by retaining the same unit of observation but changing the dependent variable" (223). (By "changing the dependent variable" DSI means changing the measure of what I call a *concept variable*: that is, a variable such as "power" or "social class" that is itself decomposable into many variables or measures. For instance, the "power" of a nation can be measured by its economic strength, its military prowess, its political cohesion, etc. I will adopt DSI's usage here and refer to a 'new measure of a DV' simply as a 'new DV'.)¹⁷⁸

Unlike the first technique, the second technique does *not* 'merely' involve stacking additional instances of an IV and DV *vertically* in an *existing* experiment. In other words, this technique does not instruct a researcher to assess the status of the regularity

¹⁷⁸ DSI creates confusion over the difference between an entirely 'new' DV and simply a new *measure* of a DV. The confusion stems from its wording in the 'social unrest' example. DSI (1994, 223) states, "In the example we are using of agricultural price fluctuation and social unrest, we may have measured social unrest by the number of public disturbances." So far, so good. But then it states, "In addition to social unrest, we might ask what else might be expected if the theory is correct. Perhaps there are other valid measures of social unrest – deviant behavior of one sort or another." These lines create confusion. One suspects that DSI was meant to read, 'In addition to *public disturbances*, we might ask what else might be expected if the theory is correct.' But we cannot be entirely certain. Regardless, I contend that adding *entirely new* IVs and DVs (not just alternative measures of a 'concept variable') can be desirable.

between the IV and *original* DV. Instead, this method of adding observations entails moving *horizontally* to one or more *new* experiments, each experiment itself yielding a regularity between the original IV and a *new* DV, and each experiment itself subject to the same obstacles discussed throughout DSI (i.e., selection bias, measurement error, endogeneity, etc.).

I use DSI's own example to illustrate the technique. DSI asks us to imagine that a preliminary empirical investigation indicates that agricultural price fluctuation is correlated with public disturbances (a measure of social unrest, which is a concept variable). Consistent with its advice of looking for observable implications of the theoretical mechanism, DSI suggests that a researcher can increase their confidence in the causal nature of this relationship by changing the dependent variable to other indicators of social unrest, including other forms of "deviant behavior" relating to "voting behavior, business investment, or emigration" (223).

There are two noteworthy points about this example and DSI's corresponding advice. First, although they never say as much, it suggests that a researcher conduct a series of *new* experiments, each one assessing the relationship between the original IV and a *new* DV (for example, attempting to correlate agricultural price fluctuation with deviant voting behavior). It might be argued that DSI had no intention of suggesting the need for a series of new experiments. But clearly testing for a correlation between agricultural price fluctuation and, say, deviant voting behaviour, would require a whole different set of control variables; for instance, factors that routinely influence voting behavior, such as

leaders' personalities. With new control variables and a new DV, it seems evident that DSI's technique suggests the need for new experiments.¹⁷⁹

Second, the new DVs suggested by DSI are *wide-ranging* and (relatively) *independent*. Public disturbances, the original DV, is itself a 'concept variable' and could be operationalized with various measures (e.g., as organized violent protests, riots, wanton violence, etc.). But DSI never suggests the researcher focus on these possibilities. Rather, they suggest that a researcher look *beyond* the 'concept variable' of public disturbances to other wider ranging measures of social unrest, such as deviant behaviour relating to voting, business investment, or emigration.¹⁸⁰ Given that DSI's technique does entail the need for a series of new experiments attempting to correlate agricultural price

¹⁷⁹ Although it is theoretically possible to include all of the DVs together in a composite measure of social unrest it not clear how this would apply to the type of small-N research that DSI addresses. Note that there is no guarantee that the unit of measurement for a new DV will be compatible with the unit of measurement for the original DV: a fact especially germane to qualitative data, the primary type of data discussed in DSI.

¹⁸⁰ DSI (1994, 222-23) explicitly discusses the importance of obtaining *independent* observations, but only in discussing their first technique for increasing the N of a small-N study, and then only in relation to the independence of individual *units*. But DSI does seem to implicitly stress independence here: after all, it is doubtful that the diverse array of indicators of social unrest they discuss – relating to voting behaviour, business investment, and emigration – would score very high on a statistical *reliability test*: that is, these measures would not correlate highly with one another.

fluctuation with a selection of wide-ranging DVs, what does this tell us? It tells us that DSI implicitly champions the importance of theoretical mechanisms.

Consider the reasoning process involved with DSI's technique. We begin with a theory that agricultural price fluctuation causes social unrest. We conduct an experiment and find that agricultural price fluctuation correlates with public disturbances, a measure of social unrest. We then ask, "If the correlation between agricultural price fluctuation and public disturbances is causal, what else should we expect to find?" To answer this question, we *must consider the mechanism hypothesized to connect agricultural price fluctuation to public disturbances*. As DSI (223) notes, "The same process that leads price fluctuations to engender unrest [by which they mean public disturbances] might link price fluctuations to these other outcomes." In the example at hand, we hypothesize (as DSI suggests) that the psychological mechanism is uncertainty (presumably economic uncertainty). That is, agricultural price fluctuation creates economic uncertainty in the population, which in turn engenders public disturbances.

Why must we consider the *mechanism* to derive further observable implications? Why not simply select any measure of social unrest and attempt to correlate it with agricultural price fluctuation? Because *there are measures of social unrest that we would not expect to correlate with agricultural price fluctuation because they do not follow from the mechanism of economic uncertainty*. For instance, widespread attacks by radical environmentalists on animal testing facilities (sometimes called ecological terrorism) could certainly be considered a valid measure of social unrest, but it does not generally

stem from economic uncertainty. We would *not* expect to find a correlation between agricultural price fluctuation and ecological terrorism.

Third Technique: New Measures, New Units

Consider now DSI's third technique for increasing the N of a small-N study: "We may also look beyond the set of explanatory and dependent variables that have been applied to a particular set of units to other observable implications involving new variables and new units" (224). Thus DSI suggests that the theoretical mechanism linking agricultural price fluctuation to social unrest within *nations* should have observable implications "in other units such as firms or government agencies" (224). For instance, we might theorize that within the agricultural industry, price fluctuations would cause firms to turn to black market activity (a possible measure of social unrest for the industry).

Each of the points I made above with respect to DSI's second technique applies with even greater force to this third technique. First, it is difficult to imagine that this technique does not entail adding one or more new experiments to an ongoing 'study'. Again, an experiment correlating price fluctuation and behaviour on the part of firms within a particular industry would require a very different set of controls than an experiment attempting to correlate agricultural price fluctuation and public disturbances within nations. As with the second technique – but unlike the first technique – this third technique does *not* 'merely' involve stacking additional instances of the original IV and DV *vertically* in an *existing* experiment but rather entails moving *horizontally* to one or more *new* experiments with a new IV and/or DV.

Second, the new experiments will entail moving to new empirical domains. As DSI (1994, 224-25) notes: “The measures used to test what are essentially new hypotheses that are derived from the original ones may be quite different from those used thus far...The movement to a new kind of “observation” – a different kind of social unit, an individual, a decision – may involve the introduction of explanatory variables [IVs] not applicable to the original unit.”

Finally, this third technique puts much greater stock in the concept of mechanism than DSI explicitly allows in its formal definition of causation. Again, we need to consider the mechanism of economic uncertainty in order to appropriately operationalize the ‘concept variable’, social unrest. This is because there are likely to be measures of “unrest” on the part of firms in an industry that do not follow from the mechanism of economic uncertainty: unless firms anticipate the complete demise of their industry, for instance, they would be unlikely to choose actions that only serve to increase price variability.

iii. Theoretical Mechanisms as per Instrumentalism Empiricism or Scientific Realism?

To what extent is DSI’s emphasis on theoretical mechanisms consistent with SR and to what extent is it consistent with IE? Unfortunately, no clear-cut answer can be provided to this question: DSI’s emphasis on theoretical mechanisms is decidedly ambiguous.

As a basis of comparison, consider again Hull, who, we have seen, embraced IE. Hull did not attribute independent causal status to theoretical entities and processes.

Unobservables were neither true nor false. They were only more or less empirically adequate. As Mills (1998, 85) points out, Hull defined intervening variables solely in terms of empirical phenomena. Thyer (1999, 5) makes a similar point: “As Hull himself put it, [theoretical constructs] ‘anchor’ their intervening variables at both their ends to empirical observations.” Hull also emphasized parsimony and generalizability. His overarching goal was to establish a compact theory portable to a wide variety of empirical phenomena. Hull’s *A Behavior System* stipulated just sixteen postulates that formed his core theory. Thus, in Hull’s work, therefore, we find the hallmarks of IE: a parsimonious set of theoretical entities and processes that is used to predict and retrodict a wide array of empirical phenomena, but which in and of itself is neither true nor false and possesses no causal efficacy.

Now consider the ambiguity of DSI. Certain passages in DSI would lead us to the belief that DSI espouses SR. First, DSI appears to hold that theoretical mechanisms are *more* than just useful fictions from which observable implications are derived. Take, for instance, a passage from their discussion of the third technique, discussed above, which emphasizes the importance of mechanism:

The general hypothesis about the link between agricultural prices and unrest may suggested hypotheses about *uncertainty* and unrest...What might we expect in terms of individual psychological *reactions to uncertainty* and the *impact of such psychological states* on individual deviant behavior?...For instance, we might hypothesize that those who are most

vulnerable to the effects of price fluctuation – growers of particular crops or people dependent on low agricultural prices for adequate food supply – would be more likely to engage in socially disruptive behavior. A test of such a hypothesis might *involve measures of psychological states* such as alienation or measures of individual deviant behavior. (224-26, my emphases)

Notice DSI's emphasis on the mechanism of uncertainty. The emphasis is no longer on whether agricultural price fluctuation is correlated with social unrest, *but rather with whether psychological states such as uncertainty and alienation are correlated with social unrest*. Moreover, DSI makes 'uncertainty' *real* by suggesting that it is a "psychological state." In other words, uncertainty appears to be real not just a useful fiction.¹⁸¹

One further comment relating to unobservable entities and processes adds evidence that DSI espouses SR. DSI states: "Inference is the process of using the facts we know to learn about the *facts* we do not know" (48, my emphasis). By referring to the 'unobservables' as "facts", DSI establishes its belief in their actual existence: it believes them to be literally real.

In addition to its implicit belief in the reality of theoretical mechanisms, DSI also explicitly argues for causation. Consider two excerpts from its philosophical discussion of causation:

¹⁸¹ This leaves open the question of whether all psychological states reduce to physical brain states.

Many social scientists are uncomfortable with causal inference. They are so wary of the warning that ‘correlation is not causation’ that they will not state causal hypotheses or draw causal inferences, referring to their research as ‘studying association not causation’...Avoiding causal language when causality is the real subject of investigation either renders the research irrelevant or permits it to remain undisciplined by the rules of scientific inference. Our uncertainty about causal inference will never be eliminated. But this uncertainty should not suggest that we avoid attempts at causal inference. (75-6)

In view of some social scientists’ preferences for explanation over “mere descriptions,” it is not surprising that students of complicated events seek to dress their work in the trappings of explanatory jargon; otherwise, they fear being regarded as doing inferior work. At its core, real explanation is always based on causal inferences. We regard arguments in the literature about “noncausal explanation” as confusing terminology; in virtually all cases, these arguments are really about causal explanation or else are internally inconsistent. (75)

These are hardly the type of comments that one would expect to hear from a HE or an IE: both would argue that that scientific inference consists of precisely what DSI explicitly rejects: namely, “noncausal explanations.”

Finally, DSI’s position on the desired complexity of theoretical mechanisms is not consistent with what one would expect to hear from a IE, although it may fall short of what of what one might expect to hear from a SR. Rather than emphasizing simple, elegant theoretical mechanisms as does IE, DSI contends that theoretical mechanisms

“should explain as much as possible with as little as possible” (29), and even goes so far as to say that “theory should be just as complicated as all our evidence suggests” (20).

If DSI’s apparent belief in the reality of theoretical mechanisms and causation leads one to conclude that it espouses SR, at least two related pieces of evidence lead us to assume that in fact its emphasis on theoretical mechanisms is consistent with IE. First, and most importantly, DSI (86-87) explicitly argues that mechanisms are useful *only* insofar as they are devices for generating hypotheses. Causal effect is “logically prior” to causal mechanism (86). Second, DSI downplays the role of case studies: it rigidly adheres to the strict demarcation between the development and testing of theoretical mechanisms.

In brief, it is entirely ambiguous as to whether DSI’s implicit emphasis on theoretical mechanisms is consistent with SR or IE. In addition, its emphasis on theoretical mechanisms – whether consistent with SR or IE – is completely inconsistent with the HE of its formal definition of causation.

iv. Conclusions: Tension in Political Science

DSI manifests a deep tension that is widespread in the discipline of political science. Like most of political science, it wants to make *causal* inference. But again, as with most mainstream political science, it does not espouse a philosophical position that complements this desire: its formal definition of causation is based in HE, and the philosophical underpinning of its emphasis on theoretical mechanisms is ambiguous. If DSI represents the last word in mainstream political science methodological prescriptions

– and the popularity of DSI would suggest that it does – then the discipline of political science is in a partial state of philosophical disarray.

Chapter XII: Causal Inference in Democratic Peace Research

Scientific realists in the social sciences typically claim that HE characterizes mainstream political science. By closely analyzing DSI, I have shown that this claim is only marginally accurate insofar as methodological prescriptions in political science are concerned. In fact, methodological prescriptions in the discipline exhibit an uneasy mix of HE, IE, and SR.

In this chapter, I analyze the democratic peace literature to demonstrate that the characterization made by most SRs in the social sciences is even further off the mark with regard to political science research. The democratic peace hypothesis holds, generally, that democracies do not fight one another and/or are inherently pacific. The democratic peace literature is widely considered among the most mature, sophisticated bodies of literature in the discipline of political science. It therefore serves as a fair ground on which to judge the philosophical basis of “state of the art” mainstream political science research.

I show that political science research stands on the cusp of SR and IE. As SR would advocate, scholars engaged in the democratic peace debate typically emphasize theoretical mechanisms as well as regularities. But in accordance with IE, these same scholars typically stress parsimony over reality in their theories, ignore the contribution of case studies, and implicitly or explicitly eschew SR-based rational choice modeling.

In the first section of this chapter, I provide a thumbnail sketch of the democratic peace literature. In the second section, I show that theoretical mechanisms play an integral role in democratic peace research. In the third section, I argue that this emphasis on theoretical mechanisms is ambiguous: some evidence suggests that it is consistent with IE, while other evidence suggests that it is consistent with SR. In the conclusions, I suggest that, compared with what SRs typically imply, relatively minor changes are required to fully implement an SR-based approach to causal inference in the discipline.

i. The Democratic Peace Debate

The following sections analyze *methodological* approaches employed by scholars engaged in the democratic peace debate. As such, my analysis does not necessarily concern itself with taking sides in the *substantive* issues debated in this literature. I include a sketch of the debate's key substantive issues for those unfamiliar with this body of literature.

Over the past two decades, scores of articles and books have been written on the democratic peace. In the modern era, the origin of the democratic peace hypothesis is widely attributed to work conducted by Melvin Small and J. David Singer at the Correlates of War (COW) project, in particular their seminal 1976 article, "The War

Proneness of Democratic Regimes.”¹⁸² The true origin of the democratic peace hypothesis, however, can be traced back at least to Immanuel Kant’s 1795 *Perpetual Peace*, in which he argued that international peace was dependent on a combination of democracy, economic interdependence, and international law and institutions. The democratic peace is often referred to as the Kantian peace.

The democratic peace phenomenon is widely heralded as one of the most important subjects of research in political science. It has profound implications inside and outside the academy. In the study of international relations, the democratic peace hypothesis represents one of the most serious challenges yet to the hegemony of Realism, because it posits that war and peace are determined by a *second image* or state-level factor (regime type) rather than *third image* or systems-level factors (anarchy and balance of power). Outside the academy, the US foreign policy of promoting democracy under the Clinton Administration was motivated in part by the notion that democracies are peaceful; the current Bush Administration claims to have the same view of the pacifying force of democracy.

The democratic peace literature contains two main schools of thought. The first school highlights two empirical regularities: (1) democracies rarely if ever go to war with one another; and (2) democracies are no less likely to engage in war with non-democracies.

¹⁸² According to Ray (1995, 11), the contemporary version of the democratic peace hypothesis actually traces back to a 1964 article by Dean Babst, a research scientist at the New York State Narcotic Addition Control Commission.

These two empirical regularities are referred to together as the *dyadic* proposition since pairings of nations determine the likelihood of war or peace (Jungblut and Stoll 2002, Owen 1994, Maoz and Russett 1993, Mintz and Geva 1993, Oneal and Ray 1997, Oneal et al. 1996, Ray 1995, Thompson and Tucker 1997).

The second school of thought contends that democracies are inherently pacific. That is, democracies are less likely than other regime types to wage war, regardless of whether they are interacting with democracies or non-democracies. The general empirical regularity between democracy and peace is referred to as the *monadic* proposition since it claims that we need only know the type of one regime to determine the likelihood of war or peace (Benoit 1996, Bremmer 1992, Rousseau et al. 1996, Rummel 1995, Russett and Oneal 2001).

For much of the 1980s and 1990s, the *dyadic* proposition appeared to prevail. Since the mid- to late-1990s, however, some opinion appears to have swung in favour of the monadic proposition: many scholars now contend that democracies are inherently more peaceful than non-democracies. Early literature on the democratic peace missed the monadic empirical regularity because it failed to notice that non-democracies are generally the aggressors in cases of conflict involving mixed dyads (democracies paired with non-democracies). The debate between the monadic and dyadic versions of the democratic peace, however, is far from over. Numerous democratic peace proponents maintain their belief only in the interdemocratic peace – i.e., the dyadic version.

What accounts for the democratic peace? Why is it that democracies do not fight one another? If democracies are less conflict-prone than non-democracies, why is this so? Or if democracies are in fact inherently peaceful, why is this the case? Two theories advanced to account for the democratic peace have dominated the literature: the *normative* (or *cultural*) theory and the *institutional* (or *structural*) theory.

The normative theory posits that democracy encourages the peaceful resolution of domestic conflict. Democracies assume that if their domestic political process resolves disputes in a peaceful fashion, then disputes with democratic adversaries will also be resolved peacefully because their adversaries will share the very same peaceful dispute resolution mechanisms. In short, peace between democracies is the result of shared political ideologies, values, and norms.

The institutional theory posits that democratic leaders are constrained by the high financial and human costs of war. Because democratic leaders are accountable to their citizens, leaders cannot easily commit the nation to war. Unlike non-democracies, decision-making in democracies requires a broad base of support, especially in the face of high-risk decisions. In high-risk decisions, democratic leaders are likely to be replaced by challengers if they make the wrong move, so they are apt to proceed cautiously. Thus democratic leaders are reluctant to wage war unless the costs are justified by the gains.

Proponents of the dyadic proposition have struggled with the institutional theory because it suggests that democracies should have peaceful relations with other states regardless of

their regime type. Various amendments were made to the institutional theory to account for this discrepancy. It was theorized, for instance, that because leaders of non-democratic nations can take advantage of the relatively deliberative process found in democracies, citizens of democratic nations are more likely to allow their leaders to engage in conflict with non-democracies without the same degree of public scrutiny.

The debates between proponents of the dyadic and monadic propositions, and between proponents of normative and institutional theories, continue within the democratic peace literature. Meanwhile, the democratic peace literature as a whole has also been the subject of much criticism and challenge. Levy (1988, 662) states that the democratic peace is “as close as anything we have to an empirical law in international relations.” But critics of the democratic peace beg to differ. Few critics dismiss the democratic peace regularities outright, but most suggest significant revisions based on challenges to one or more of its central empirical and/or theoretical tenets. Critics charge that the empirical regularities associated to the democratic peace may be either spurious relationships or statistical artifacts, because they are based on insufficient data, flawed definitions of democracy, flawed definitions of war, a failure to control for confounding factors, a lack of historical context, and a failure to specify a compelling theoretical mechanism (Barkawi and Laffey 2001, Elman 1997, Farber and Gowa 1997, Gartzke 1998, Gowa 1999, Layne 1994, Moore 2004, Remmer 1998, Schwartz and Skinner 2002, Snyder 2000, Spiro 1994, Rosato 2003).

ii. The Importance of Theoretical Mechanisms in Democratic Peace Research

A key issue at stake in the democratic peace debate, then, is whether or not the dyadic and monadic findings are spurious or whether they are genuinely causal. Small and Singer originally even dismissed their own findings on the democratic peace, believing “that the relationship was spurious, perhaps merely a result of democracies’ being physically far apart” (Russett and Oneal 2001, 49). Proponents of the democratic peace have responded to the potential for spuriousness with various strategies.

One response to potential spuriousness is consistent with HE. Several researchers have increased the sophistication of their statistical techniques in order to more accurately gauge the correlation between democracy and peace. Raknerud and Hegre (1997), for instance, develop and employ advanced statistical techniques in order to address some potential weaknesses of early less-sophisticated tests of the dyadic peace. Similarly, Benoit (1996) argues that the lack of support for the monadic peace is a result of impoverished statistical techniques. He applies advanced techniques that find support for the monadic peace. Rousseau et al. (1996) also introduce some statistical innovations to better assess the dyadic and monadic findings, as do Huth and Allee (2002).¹⁸³

A second response to potential spuriousness has been to control for confounding factors that are temporally prior to the key IV. For instance, Oneal and Ray (1997, 755) control

¹⁸³ This is not say that the SSR approach is opposed to advanced statistical techniques. What the SSR approach is opposed to is increasing the sophistication of statistical techniques to the exclusion of similar theoretical advancements.

for economic interdependence on the assumption that “democracy has flourished in wealthy states linked by trade and investment, making the costs of war prohibitive.” Numerous other works on the democratic peace likewise control for the confounding influence of economic interdependence (e.g., Bremmer 1992, Maoz and Russett 1993, Oneal et al. 1996).¹⁸⁴

These two approaches to addressing spuriousness are entirely reasonable: the sophistication of the problem area should be matched with equally sophisticated diagnostic tools; and the impact of confounding influences has to be assessed. But what is *most* striking about the democratic peace literature is that the much more common approach to addressing potential spuriousness is to investigate the theoretical explanations for *why* democracies do not fight one another. In other words the assumption is that if the mechanism can be firmed up, then confidence in the causal nature of the democratic peace will be greatly enhanced.

It must be said that, even with the two aforementioned strategies for dealing with possible spuriousness, theoretical investigations into mechanisms sometimes play a role. For instance, Raknerud and Hegre (1997) apply their statistical innovations to assess certain

¹⁸⁴ Scholars in the democratic peace debate also routinely control for factors that are not necessarily confounding but that might have strong causal association to patterns of war and peace, such as alliances and geographic contiguity. These factors are generally included as IVs because scholars want to assess the relative explanatory powers of (IR) realist factors against liberal factors.

theoretical innovations. More specifically, they use advanced statistical modeling to determine how likely democracies are to join one another in war – an issue that has implications for the monadic peace. (Like Gleditsch and Hegre (1997), Raknerud and Hegre (1997) find that democracies are as war-prone as non-democracies.) Huth and Allee (2002) similarly develop their statistical innovations for the express purpose of testing theoretical refinements.

For the most part, however, theoretical work has not been the byproduct of, or an afterthought to, statistical innovations. The theoretical mechanism invoked to explain the democratic peace has been, *in and of itself*, a source of copious amounts of research, and a key strategy in dispensing with potential spuriousness. Russett and Oneal (2001, 53) make this point: “A strong empirical relationship between democracy and peace alone is not enough. We need to know *why* such a relationship exists; without a theoretical explanation, we do not understand the cause of the phenomenon and cannot be sure that the finding is not purely coincidental.”

In the next section I focus on the development and testing of theoretical mechanisms in Bruce Russett and John Oneal’s 2001 *Triangulating Peace*. Russett and Oneal are widely regarded as key scholars in the democratic peace literature. *Triangulating Peace* represents the culmination of a decade’s worth of work by these scholars. As we will see, Russett and Oneal’s emphasis on theoretical mechanisms is fairly typical of research in the democratic peace literature. The following section extends the discussion to the

numerous other scholars from the democratic peace literature to demonstrate that attention to the development and testing of theoretical mechanisms is widespread.

Development and Testing of Theoretical Mechanisms in *Triangulating Peace*

SRs in the social sciences typically accuse mainstream social scientists of basing their causal finding merely on a-theoretical correlational evidence: that is, they argue that mainstream social science is characterized foremost by HE. Almost all the work conducted on the democratic peace hypothesis belies this claim. A core feature of the democratic peace literature is the development of theoretical mechanisms.

A foremost aim of Russett and Oneal (2001) is to refine and expand on traditional normative and institutional theories for the democratic peace. To this end, Russett and Oneal (24-58) draw on Kant's *Perpetual Peace*. Following Kant, they argue for the compatibility rather than competitiveness of normative and institutional theories: "...a good constitution for representative government would, over time, generate a good moral culture." And, like Kant, Russett and Oneal emphasize the role of economic interdependence and international institutions as pacifying forces. They contend that peace is the result of *triangulation* between democracy, trade, and international institutions: each of these three factors complement one other in a "virtuous circle" to generate peace. Their Kantian-based theoretical mechanism, for instance, predicts that democracies are more likely to trade with one another; and interdependence, in turn, fosters democracy. Similarly, democracies are more likely than non-democracies to join international institutions; and international institutions, in turn, promote democratization.

And international organizations are often formed with the intention of promoting economic interdependence. Thus feedback loops between all three factors are theorized to explain the democratic peace, although democracy itself remains the focal point of their liberal or Kantian peace (38).

If Russett and Oneal's focus on developing theoretical mechanisms belies the typical SR's characterization of social science as being based in HE, so too does the manner in which their theoretical mechanism is tested. Russett and Oneal test their theoretical mechanism against a range of observable phenomena, acknowledging that this method is key to scientific progress: "...[T]he basic proposition about peaceful relations among democracies serve[s] as a basis for developing theories to explain why that might be true. Generating theories and then comparing them to the evidence force[s] analysts to consider how well they [explain] not only this but other empirical phenomena."

In defending their theoretical mechanism for the democratic peace – which includes, normative elements of shared values and goals – Russett and Oneal (2001, 59) note: "If democracies reap rewards from avoiding conflict with each other because they share common interests, we should expect those common interests to show up in a wide range of cooperative behavior." Thus Russett and Oneal (59-62) point to various observable implications of their theoretical mechanism of "shared interests." For instance, they point to evidence for the willingness of democracies "to collaborate with each other at the start of militarized disputes." Contra what IR realists would predict, they also point to evidence that non-militarized disputes between democracies are settled by compromise

reached through negotiated settlement, even in cases in which the power of the democratic party to the dispute far outweighs the other. Non-militarized disputes between non-democracies or mixed dyads, conversely, are usually settled in favour of the more powerful nation by force or persuasion. Russett and Oneal (64-5) further point to evidence that “[d]isputes between democracies are both shorter (nearly half lasted only a day) and less severe than those between other kinds of states,” and that “[d]emocracies are much more likely to conclude preferential trade agreements.”¹⁸⁵ In short, Russett and Oneal test their explanation for the democratic peace by moving horizontally to a range of new empirical domains.

Russett and Oneal’s explanation of the democratic peace contains institutional as well as normative elements: the lack of war between democracies is not just the result of shared interests, but also the result of the democratic process that penalizes costly decisions. To demonstrate the validity of the institutional theory, Russett and Oneal (66-7) point to evidence that democracies generally win the wars that they fight, and suffer fewer casualties than do autocracies. The explanation for these additional empirical regularities, Russett and Oneal argue, is the very same explanation for why democracies do not fight one another: accountability makes democratic leaders chary in regard to the potentially high economic and human costs of war. Democracies are careful to pick only

¹⁸⁵ Russett and Oneal fail to acknowledge that some of this evidence works against their monadic proposition even as it supports their dyadic proposition. But what matters for my purposes is not whether their argument is compelling but rather their methodology for attempting to secure their argument.

those fights that they have a good chance of winning, and, unlike autocracies, choose military strategies that minimize the risk of casualties when they do choose to fight. Thus again, Russett and Oneal move horizontally to new empirical domains to test their theoretical mechanism.

Russett and Oneal (68-70) also move horizontally by disputing empirical evidence that appears to contradict institutional elements of their theoretical mechanism. They note that many political commentators believe that democratic leaders often bring their nations to war as a diversionary tactic: the so-called “rally ‘round the flag” phenomenon. When Bill Clinton ordered the bombing of suspected terrorist sites in Sudan and Afghanistan in 1998, pundits accused the President of merely attempting to divert attention from the Monica Lewinsky scandal. If true, the “rally ‘round the flag” phenomenon would undermine the institutional theory for the democratic peace because it would suggest that democratic leaders were not seriously constrained by public pressure in their ability to wage war. But Russett and Oneal point to systematic empirical studies that contradict the phenomenon. These studies “find only modest support for the existence of a ‘rally’ effect following a use of force by an American president,” that there is little evidence “...that American leaders have tried to invoke a rally at politically convenient times,” and that “[i]nternational influences usually have greater impact on American decisions to use force than domestic conditions do, especially if one looks at situations in which the US government might have used military force but did not.”

The above provide just a few examples of the manner by which Russett and Oneal (2001) test their explanation of the democratic peace by moving horizontally to new empirical domains. There are many other examples that one could discuss. They test the independent and mutual influences of their three Kantian factors – democracy, economic interdependence, and international institutions – against a variety of empirical phenomena (125-238). They also test their theoretical mechanism against empirical evidence relating to civil wars (70-1), the war-proneness of non-democratic dyads (114-16), and democracies in transition (116-22), to name a few. The point, however, should be clear – without going into further detail – that Russett and Oneal’s work on the democratic peace is not HE-based. This finding puts to lie a standard assumption of SRs in the social sciences.

Theoretical Mechanisms in the Wider Democratic Peace Literature

Evidence of attention to the development and testing of theoretical mechanisms in the democratic peace literature is hardly unique to Russett and Oneal’s work. In fact, democratic peace scholars routinely make refinements to normative and institutional theories and move horizontally to test their refined explanations against a range of observable implications.

For instance, Huth and Allee (2002, 5), state their concern for the development of their mechanisms:

[W]e think supporters are right that hypotheses about norms of political bargaining or the accountability of leaders to political opposition represent plausible and fruitful theoretical

approaches to explaining how domestic political institutions influence the foreign policy choices of state leaders. However, we believe that for both the norms-based and accountability-based approaches, the logical hypotheses to be tested are not adequately established in the existing literature. Through critical re-examination of the theoretical foundations of each approach, we can develop new hypotheses that refine and extend existing arguments.”

Huth and Allee develop their own “Political Norms Model” (101-8), identify a range of observable implications (108-17), and statistically test these hypotheses (138-276). Huth and Allee contend that as standard norms theory predicts, the monadic proposition holds: although mixed dyads are more likely than democratic dyads to engage in conflict, non-democracies are generally the aggressors. However unlike standard norms theory, Huth and Allee’s “Political Norms Model,” suggests that democratic leaders will not “attempt to ensure their security by adopting more inflexible and coercive policies when facing an adversary with non-democratic norms of bargaining” (116). Rather, democratic leaders are consistent in their application of “tit-for-tat” diplomatic and military policies, regardless of regime type. Disputes with non-democracies, however, can more easily escalate to violent conflict because non-democratic leaders “with very violent norms are more willing to bargain in a more inflexible and coercive manner” (116). Huth and Allee derive multiple hypotheses around this particular version of norms theory, and test them against a range of empirical evidence.

Bueno de Mesquita et al. (1999) are even more explicit about the need for a theoretical mechanism to explain the democratic peace that is portable to a wide range of empirical

domains. They identify several regularities associated to the democratic peace, and argue that:

...any explanation of the democratic peace must satisfy two criteria. First, it must account for the known regularities that are often grouped together to define the democratic peace...Second, we believe that a credible explanation also should suggest novel hypotheses that do not form part of the corpus of the democratic peace. Further credibility to the overall explanation is added if these novel hypotheses are borne out by the evidence. (792)

To that end, Bueno de Mesquita et al. contend that the normative and institutional theories are flawed and meet neither criterion. They develop a game theoretic model that accounts for the known regularities and derive two novel hypotheses from their game theoretic model that are supported by existing empirical evidence. (Greater detail on Bueno de Mesquita et al. (1999) is provided later in this chapter.)

Or take Lipson (2003), who develops a theoretical explanation for the democratic peace and argues:

If this explanation is correct, it should not only clarify why democracies so rarely fight each other, it should tell us much more about democracies in world politics. We should expect a series of other major empirical findings...These extensions are important for both substantive and methodological reasons. (8)

Rosato (2003) is equally explicit about the need for the development and testing of theoretical mechanisms. Rosato (2003, 585) contends:

To test a theory fully, we should determine whether there is a correlation between the independent and the dependent variables and whether there is a causal relationship between them. An evaluation of the democratic peace theory, then, rests on answering two questions. First, do the data support the claim that democracies rarely fight each other? Second, is there a compelling explanation for why this should be the case?

Rosato uses two techniques to determine whether or not the relationship between democracy and peace is causal. First, he examines the causal logics of the mechanisms suggested by normative and institutional theories. Rosato notes that “A causal logic is a statement about how an independent variable exerts a causal effect on a dependent variable. It elaborates a specific chain of causal mechanisms that connects these variables and takes the following form: A (the independent variable) causes B (the dependent variable) because A causes x, which causes y, which causes B.”

Second, Rosato uses the causal “logics to generate additional testable propositions about the effects of democracy on state behavior. If we accept that A does cause x, that x causes y, and y causes B, then logical deduction can yield other propositions that should also be true. These too can be checked against the historical record, and the theory will be strengthened or weakened to the extent that they find empirical support” (586).

Rosato's analysis suggests that the causal logics of the normative and institutional theories are flawed and their testable propositions are not borne out by the historical record. He concludes therefore that, "there are good reasons to believe that while there is certainly peace among democracies, it may not be caused by the democratic nature of those states" (585). In short, Rosato concludes that because the mechanisms alleged to connect democracy to peace is unsound, the correlation may be spurious.

Even when the development of theoretical mechanisms is not the focus of democratic peace research, scholars recognize that empirical regularities cannot stand alone. Based on empirical evidence, Rousseau et al. (1996) argue that the democratic peace contains mixed dyadic and monadic effects. In times of crisis, democracies will escalate to conflict against non-democracies but not against fellow democracies, a phenomenon that supports the dyadic proposition. In non-crisis periods, however, democracies are unlikely to initiate crises or conflicts with democracies or non-democracies, a phenomenon that supports the monadic proposition. In their concluding remarks, Rousseau et al. (527-8) briefly sketch a possible theoretical mechanism to explain these empirical findings, and emphasize that future empirical research should aim to further test this proposed mechanism. In other words, Rousseau et al. recognize that their *ad hoc* theoretical mechanism cannot be used to explain their novel empirical regularities – regardless of the fact that they are robust – without further testing.

I could point to many more examples of attention to the development and testing of theoretical mechanisms in the democratic peace literature. These are at least partial focal

points for Jungblut and Stoll (2002), Mintz and Geva (1993), Oneal et al. (1996), Raknerud and Hegre (1997), and Starr (1997), just to name a few. Individual democratic peace studies are not necessarily comprehensive in their development and testing of theoretical mechanisms: practical impediments – funding restrictions and space limitations, for instance – necessarily put limits on what scholars can achieve in individual studies. But it is worth noting that *collectively*, the democratic peace literature features incremental progress in developing and testing theoretical mechanisms.

iii. Scientific Realism or Instrumentalism Empiricism in the Democratic Peace Literature?

The manner by which democratic peace scholars pay attention to the development of mechanisms, and the manner by which they move horizontally to test these mechanisms against a range of empirical domains, confirms that their approach is not consistent with HE. But it leaves open the question as to whether their approach is more consistent with SR or IE.

Recall from Chapter X that both SR and IE adopt the strategy of testing theories against a range of empirical phenomena. The crucial difference between these two philosophical positions of thought pertains to the trade-off between generalizability and accuracy, and by extension, to the conclusions that one can draw about the truth or falsehood of theoretical claims. IE favours parsimonious theories and simplifying assumptions that are portable to a wide variety of empirical domains. For IE, empirical affirmations do not provide grounds for the literal truth or falsehood of theoretical claims. Theories are simply useful heuristic devices for prediction. SR too favours portability to an array or

empirical domains, but is willing to sacrifice a measure of generalizability on the altar of theoretical accuracy. Empirical confirmation or disconfirmation of theoretically accurate mechanisms that attempt to model reality, provide, at least in part, grounds for belief in the literal truth or falsehood of claims about theoretical mechanisms.

Where does the democratic peace literature come down on the tension between theoretical parsimony and accuracy? The answer is not clear-cut one way or the other. Some pieces of evidence seem to suggest that democratic peace scholars favour IE, while others would have them favour SR. The literature seems to sit on the cusp of both philosophical positions and is often inconsistent.

The inconsistencies are most evident in regard to physicalness and intentionality – respectively, the first and third ontological properties of social mechanisms – and their corresponding methods of process-tracing and interpretation. Elman (1997, 33) – at the time, one of the few democratic peace scholars to champion process-tracing – argues that “[I]f we want to move beyond correlation to causation, we need to reveal the decision making processes of aggressive and pacific states” through “...detailed historical analysis...” But leading democratic peace scholars reject or seriously downplay the value of case studies, even as they use selective real-world episodes to support their own theories. Russett and Oneal (2001, 47), for instance, reject outright Elman’s compilation

of case studies. Huth and Allee (2002, 6) are only slightly more generous toward case study research in general.¹⁸⁶

The rejection or downplaying of case studies is somewhat mystifying, given that democratic peace scholars routinely draw liberally from various real-world examples in order to formulate their own mechanisms, something that Elman (1997) points out as well. Moreover, based on their dedication to the refinement of the normative and institutional theories, democratic peace scholars would appear to favour progressively more complex mechanisms. Presumably, democratic peace scholars draw on real-world examples and favour theoretical sophistication because they favour historical accuracy and realism in their models. Given that case studies make serious efforts at drawing out these historical lessons in all their complexity, their rejection of case study methods seems unwarranted, and inconsistent with their actual use of real-world data.¹⁸⁷

¹⁸⁶ My reading of the democratic peace literature on this point differs from George and Bennett (2005, 46) who contend: “Both proponents and critics of the existence of a democratic peace agree on the importance of process-tracing on causal mechanisms, and researchers who had once relied largely on statistical methods have turned to case study methods because of these methods’ ability to test causal mechanisms.”

¹⁸⁷ Granted, both Russett and Oneal (2001) and Huth and Allee (2002) reject case studies as a definitive method of *testing* the democratic peace. But as argued in Chapter X, the strict demarcation between the testing and development of mechanisms is questionable. Almost all the case studies contained in Elman (1997) derive new theoretical propositions, regardless of how they fare as tests of the democratic peace. But Russett and Oneal (2001) and Huth and Allee (2002) fail to acknowledge the contribution that these and other case studies make toward the

The democratic peace literature appears equally inconsistent with regard to rational choice-based mechanisms. This is evident, for instance, in rational choice research conducted by Bruce Bueno de Mesquita on the democratic peace. As noted earlier, Bueno de Mesquita et al. (1999) argue explicitly that the democratic peace requires a theoretical mechanism that can stand up to empirical tests against a range of empirical phenomena. What is less clear from this work (and earlier works by Bueno de Mesquita on the democratic peace) is whether he advocates IE or SR.

Recall from Chapter X that a SR-based approach to rational choice modeling can be distinguished from an IE-based approach based on differences as to how each approaches rationality, self-interest, and methodological individualism. Generally, with regard to these three issue-areas, SR champions realistic rational choice models that emphasize accuracy over parsimony and generalizability, whereas IE emphasizes empirically adequate rational choice models that favour parsimony and generalizability over theoretical accuracy.

development of theoretical mechanisms. The opinions of these leading democratic peace scholars are fairly representative of the field at large: With some notable exceptions (e.g., Ray 1995, 131-57) most democratic peace scholars are largely silent on the value of case studies, a silence that speaks to the lasting influence of IE (and, to a lesser extent, HE): case study research is typically viewed as a separate, distinct, and inferior methodology when compared to correlational research.

Bueno de Mesquita (1999, 792) explicitly acknowledges that parsimony is paramount in the development of theoretical mechanisms, and so appears to place himself squarely within the camp of IE.¹⁸⁸ But his approach to rationality, self-interest, and methodological individualism is not always consistent with IE.

First, consider the issue of rationality. Recall from the previous chapter that SR favours a “domain response” that emphasizes the reality of a model, whereas IE favours an “as if response” that emphasizes empirical adequacy. In his review of the democratic peace literature, Ray (1995) points out that Bueno de Mesquita is inconsistent on this issue-area. Bueno de Mesquita (quoted in Ray 1995, 134) contends: “I do not suggest that decision-makers *consciously* make the calculations of the expected utility model. Rather, I argue that the leaders act *as if* they do.” But Ray (1995, 134) points out that Bueno de Mesquita’s overt IE is belied by an implied SR:

Mesquita rather consistently attempts to demonstrate (in a confusing, if not contradictory, way) that in fact decision makers *do* make calculations of the type incorporated into his expected utility theory. For example, in the same article in which he disavows any claim that decision makers consciously make the calculations in his expected utility model, he

¹⁸⁸ Bueno de Mesquita creates some confusion on his position, however. He argues that even if a *less* parsimonious theory could explain more regularities, the more parsimonious theory is to be preferred: “...the more patterns that are explained, the more credible the explanation, provided that it does not come at the expense of parsimony.” As explained earlier, however, it is usually the case that parsimony in theory goes hand-in-hand with generalizability, not vice versa as Bueno de Mesquita seems to imply.

also argues that ‘the concepts of equation [5] can be seen at work in the decision of Istvan Tisza, the Magyar premier, in July 1914, to endorse military action against Serbia (Bueno de Mesquita 1988, 643). Similarly, in *War and Reason*, Bueno de Mesquita and his coauthor (Bueno de Mesquita and Lalman 1992, 225-43) seem obviously quite intent on demonstrating that the calculation of Bismarck and other decision makers involved in the Seven Weeks’ War in 1866 between Prussia and Austria paralleled quite closely those stipulated by the international interaction model developed in the book. Readers might well wonder, with all the emphasis that these models put (in typical formal modeling fashion) on the ‘as if’ assumption, why they should emphasize precisely how decision makers think and calculate in the crises being analyzed.

Ray (1995, 144) concludes that one can certainly be excused for concluding on the basis of an explicit epistemological position that rational choice modelers such as Bueno de Mesquita are interested solely in regularities and not in the causal processes that engender those regularities. But as Ray points out, Bueno de Mesquita’s rational choice modeling of the democratic peace hypothesis reveals a conflict because it does, in fact, attempt to model real-world processes.

Next, consider the issue-area of self-interest. Again, recall from Chapter X that SR favours a “thin subjectivist” response that restricts generalizability by allowing actors to be motivated by factors other than self-interest, whereas an IE favours a “thick objectivist” response that restricts actors’ motivations to self-interest. On this issue-area, Bueno de Mesquita’s model is consistent with his overt emphasis on parsimony (and, therefore, IE): “Our model explains...diverse phenomena without attributing superior

motives or greater civic mindedness to one kind of leader over another. The explanation is driven purely by self-interested leaders who seek to retain office and face alternative institutional arrangements” (805).

Finally, consider the issue-area of methodological individualism. SR holds that rational choice modeling can incorporate structural position and the constitution of actors into its purview, whereas IE rejects the notion that “rules of the game” can be introduced exogenously into its models. The model developed by Bueno de Mesquita et al. (1999) is based on the notion that leaders in democracies and non-democracies “face alternative institutional arrangements” (805). Contra what IE advocates, the “rules of the game” do not stem solely from the interactions of individuals but rather from differing institutional contexts.

In sum, rational choice models developed by Bueno de Mesquita – a leading democratic peace scholar – contain elements of both SR and IE. Bueno de Mesquita advocates parsimony above all else in the development of theoretical mechanisms, an indication of IE. His approach to self-interest is also consistent with IE: he espouses “thick rationality.” However, his approach to methodological individualism is more consistent with SR than IE as he appears to believe that “rules of the game” can be introduced exogenously. And although in his approach to rationality, Bueno de Mesquita favours IE (because he explicitly advocates an “as if” response), Ray (1995) points out that in practice his work is more consistent with SR (because he implicitly advocates a “domain response”). I conclude that as with physicalness and intentionality, democratic peace

research that relates to agency – the second ontological property of social mechanisms – and its corresponding method of rational choice based modeling, is often inconsistent.

A final area of inconsistency relates to policy implications that nearly all democratic peace scholars draw from their respective studies. Theoretical propositions are routinely transformed into policy prescriptions, something that implies tacit ontological realism: if the theoretical propositions were merely useful fictions, how could they possibly form a basis for policy? Huth and Allee (2002, 18), for instance, note that past democratic peace theory suggests that democratic leaders “may find it difficult to credibly signal their resolve in military confrontations, or to induce concessions from authoritarian leaders in negotiations.” “One policy implication that follows from such expectations,” argue Huth and Allee, “is that democratic leaders should have limited confidence in their ability to bargain effectively with authoritarian states in crises during peacetime negotiations.” By contrast, as discussed above, Huth and Allee’s “Political Norms Model” (116) suggests that democratic leaders “do not view themselves as vulnerable,” which suggests a different set of policy prescriptions for these leaders in dealing with non-democracies. Regardless of which theoretical mechanism should prevail, the point is that Huth and Allee turn theoretical propositions about mechanisms into policy prescriptions. This is common to the democratic peace literature and suggests tacit SR.

iv. Conclusions: Methodological Prescriptions and Actual Research

The findings in this chapter suggest that political science research is not as far from SR as most SRs in the social sciences imply. However, as with the discipline’s methodological

prescriptions, actual research in the discipline does exhibit philosophical inconsistencies. In the following chapter, I use the material on DSI developed in Chapter XI and material on the democratic peace hypothesis developed in this chapter to demonstrate how these philosophical inconsistencies in the disciplines' methodological prescriptions and actual research can be resolved in favour of SR. I suggest that by fully adopting the SSR approach to causal inference, the discipline can be advanced on several fronts.

Chapter XIII: The SSR Approach to Causal Inference as a Basis for Progress in Political Science

So far in Part III of this thesis, I have examined the philosophical foundations of methodological prescriptions and actual research in political science. The purpose of the examination is to try to determine where the discipline of political science sits vis-à-vis three competing philosophical visions of causal inference: HE, IE, and SR. Typically, SRs in the social sciences claim that HE characterizes the discipline of political science. My findings only modestly support this characterization. Instead, my findings indicate that, in its methodological prescriptions, political science exhibits an inconsistent mix of HE, IE, and SR; and in its actual research, political science exhibits an uneasy mix of IE and SR.

In sum, I find that the discipline of political science *has* no firm philosophical foundation. Its methodological prescriptions are self-contradictory as are its actual research practices, and the former often do not mesh with the latter. If political scientists are serious about their desire to make genuine causal inferences, then it makes good sense to adopt an SR-based approach: neither HE nor IE provide philosophical grounds on which to base causal research. But unlike what typical SRs imply, I argue that the discipline does not require a radical rethinking in order to adopt SR.

In this chapter, I want to suggest that by adopting the SSR approach to causation, political science will not only achieve philosophical consistency, but might also make

progress on at least three fronts: the SSR approach promises to produce better-trained practitioners, enhance epistemic unity in the discipline, and improve pragmatic solutions for policymakers. Throughout this chapter, I provide specific examples from DSI and the democratic peace literature, in order to demonstrate how practitioners can more fully adopt the SSR approach to making causal inference.

The first section of this chapter explains how the SSR approach can yield better-trained practitioners by “making sense” of, and even extending some of, DSI’s useful pragmatic research techniques. The second section demonstrates how the SSR approach can help bridge several methodological divides in the discipline, including nomothetic-ideographic divides (between statistical analysis and case-study research, and between rational choice modeling and case-study research), a nomothetic-nomothetic divide (between statistical analysis and rational choice modeling), and an ideographic-ideographic divide (between process-tracing and interpretation). The third section argues that the SSR approach can yield better advice to policymakers because it emphasizes detailed and accurate mechanisms. The conclusions remind readers of the “big payoff” that the SSR approach might provide: namely, a philosophical foundation and methodological toolkit that complements the desire on the part of political scientists to make *causal* inference.

i. Training Practitioners

Although diversity of philosophical traditions is in some sense welcome in political science, it becomes problematic when the diversity manifests itself as contradictory views

within a single text. As argued, such is the case with DSI. DSI's hodgepodge of HE, IE, and SR can only create confusion for political science practitioners.

For instance, consider again the difference between DSI's first technique for increasing the number of observations of a small-N study and their second and third techniques.

The first technique involves adding observations to an existing experiment, whereas the second and third techniques amount to adding experiments to an existing study. DSI never discusses this distinction, arguing simply that all three techniques can be used to increase the number of observations of a small-N study. Moreover, DSI's formal definition of causation cannot make sense of the second and third techniques: moving horizontally to new empirical domains does not mesh with the HE of its formal definition of causation.

These methodological and philosophical problems with DSI might create confusion, but do they really matter? Can practitioners not ignore these methodological and philosophical problems and simply follow DSI's advice for making causal inference? I contend that simply following rules of causal inference without an understanding of their philosophical underpinning can lead to problems. To continue with the example at hand, without an understanding of the distinction between DSI's three techniques for "making many observations from few," a practitioner might rest content with using only the first technique. Based on DSI's lack of distinction between the three techniques, a practitioner might reasonably assume that boosting one's N with first technique alone is sufficient. The SSR approach, however, requires the second and/or third techniques in

order to make causal inference: it is not enough to stack observations vertically in an existing experiment; causal inference requires moving horizontally to new empirical domains.

Blindly following methodological advice without understanding its philosophical underpinnings can also lead practitioners to miss additional techniques for increasing the number of observations of a study. With its second technique DSI suggests "...looking for many effects of the same cause...by retaining the same unit of observation but changing the dependent variable" (223). SR (and IE), however, suggests that it is also quite plausible to look for 'different causes of the same effect' by retaining the same unit of observation but changing the *independent* variable (to a new measure of the 'concept IV' or even to an *entirely new* IV). Consider again DSI's example of agricultural price fluctuation and social unrest. In order to test the relationship between uncertainty (the mechanism) and unrest (the DV), a researcher could look for IVs unrelated to agricultural price fluctuation that might engender uncertainty (and, presumably, social unrest). For instance, a researcher may want to attempt correlating social unrest with the occurrence of government or corporate scandals, or an energy shock, in order to test the relationship between economic uncertainty and social unrest.

Controlling for the Impact of the Key IV

A key rule that DSI adopts for making causal inference is a testament to the influence of HE. DSI argues that one should never control for variables that are the consequence of the key IV. Thus, in their hypothetical "incumbency" example (see Chapter XI for

details on this example), DSI (78) warns that we should “*not* control for qualities of the candidates, such as name recognition, visibility, and so forth. If we did hold these constant, we would be controlling for and hence disregarding some of the most important effects on the vote total. In fact, controlling for enough of the consequences of incumbency could make one incorrectly believe that incumbency had no effect at all.”

The reasoning implicit in DSI’s advice is that there could be multiple and complex paths leading from incumbency to vote totals. Figuring out which, if any, of these mechanisms are actually at work, and in what proportion relative to one another, is not necessary. All that *is* necessary is that the researcher ascertains *whether or not* a causal effect (i.e., a correlation) exists between incumbency and vote totals.

The SSR approach to causal inference, by contrast, contends that selectively controlling for intervening variables is an effective way of determining what mechanisms are at work, and how important these mechanisms are relative to one another.

In explaining their own SR approach, Pawson and Tilley (1997) make this point with an example from their field of evaluation research. They note that in an effort to determine whether closed-circuit television (CCTV) has a causal impact on crime rates in parking lots, the Humean would ‘simply’ compare the crime rates before the installation of CCTV

in a series of parking lots with the crime rates after the installation of CCTV. But Pawson and Tilley (1997, 78) argue that SR takes a different approach¹⁸⁹:

...[T]here is nothing about CCTV in car parks which intrinsically inhibits car crime. Whilst it may appear to offer a technical solution, CCTV certainly does not create a physical barrier making cars impenetrable. A moment's thought has us realize, therefore, that the cameras must work by instigating a chain of reasoning and reaction. Realistic evaluation is all about turning this moment's thought into a comprehensive theory of the mechanisms through which CCTV may enter the potential criminal's mind...

Pawson and Tilley (1997, 79-80) go on to suggest some ideas for tailoring correlational tests to the hypothesized mechanism(s), including systematically controlling for possible intervening variables.

Consider as well some democratic peace scholarship from Maoz and Russett (1993), who attempt to control for the potential influence of economic interdependence. Maoz and Russett recognize that economic interdependence might not be a *confounding* factor (one that precedes and accounts for both democracy and peace), but rather an *intervening* factor (whereby democracy induces economic trade, which, in turn, prevents conflict by making the cost of war prohibitive). In controlling for economic interdependence, therefore, Maoz and Russett (1993) directly contravene DSI's advice to never control for consequences of the key IV. In part, Maoz and Russett (1993) control for economic

¹⁸⁹ Pawson and Tilley champion critical realism, which as explained in Chapter III, is closely related to SR.

interdependence because they want to assess the extent to which economic interdependence is the mechanism at work in the relationship between democracy and peace.

ii. Epistemic Unity

In addition to producing better-trained practitioners, the SSR approach can also bring a measure of epistemic unity to a discipline deeply divided along at least three major fault lines.¹⁹⁰ First, qualitative researchers, who favour ideographic methods such as interpretation and process-tracing, are often at odds with quantitative researchers who favour nomothetic methods such as statistical analysis and rational choice modeling. Peattie (1995, 394) argues that in urban research, statistical methods and case study methods are generally “different in structure and are practiced by rather different sets of professionals.” In fall 2000, hundreds of tenured political science professors signed a letter protesting policies and practices at the American Political Science Association (APSA), which they accused of emphasizing quantitative approaches such as statistical analysis and rational choice modeling to the near exclusion of qualitative methods such as process-tracing and interpretation (Bennett 2002, Kasza 2001).¹⁹¹

¹⁹⁰ In the social sciences, the methodological divide is often overly simplified as one strictly between quantitative and qualitative research. For example, see Bryman (1988).

¹⁹¹ This protest letter turned into a full-blown protest movement referred to as the “Perestroika movement.” The movement has met with some success in achieving its goals. For instance, the APSA started a journal, *Perspectives on Politics*, which publishes a broader range of articles in terms of substance and methodology than does its flagship journal, the *American Political Science Review*. Nevertheless, fundamental quantitative-qualitative divisions persist in the discipline.

Second, amongst quantitative researchers who favour nomothetic methods, there is sometimes debate between proponents of statistical analysis and proponents of rational choice modeling. Although both types of methodologies are highly quantitative, there are deep philosophical divisions over the nature and testing of theory (Friedman 1995, Green and Shapiro 1994).

Third, amongst qualitative researchers that favour ideographic methods, there is debate between proponents of process-tracing and proponents of interpretation. The general view here is that tracing causal linkages is a fundamentally different type of activity from understanding intersubjective meanings.

Statistical Methods vs. Process-tracing, Interpretation, and Rational Choice

As argued, the SSR approach holds that mechanisms and regularities are integral to causation. Mechanisms can be identified by the methods of process-tracing, rational choice, and interpretation; and, according to the logic of the AfC, causal inference is made by testing these mechanisms against a range of regularities that are determined by

Rather than *separate* journals for quantitative and qualitative research, the SSR approach suggests that qualitative and quantitative research should be *integrated* in the very *same* journals. George and Bennett (2005, 3-5) provide evidence to suggest that the discipline of political science has, in fact, become less divisive since the mid-1980s than it was in the 1960s and 1970s. However, they also report that since the mid-1980s, only one in five articles in the top political science journals in the discipline use more than one method.

statistical methods.¹⁹² At the same time, methods that identify mechanisms can serve as provisional tests of regularities: if mechanism-identifying methods (process-tracing, rational choice, and interpretation) cannot reveal an explanation for a regularity that is based in empirical reality, the causal nature of that regularity is (or should be) called into question. It follows, therefore, that for the SSR approach, nomothetic statistical methods are complementary to – rather than in competition with – ideographic methods (process-tracing and interpretation) as well as other nomothetic methods (rational choice). This complementarity addresses, in part, the first aforementioned disciplinary divide between ideographic and nomothetic methods (I will discuss shortly complementarity between rational choice and ideographic methods), as well as the second aforementioned disciplinary divide between proponents of nomothetic methods.¹⁹³

One might assert that IE provides a basis for complementarity in the discipline on the same basis as the SSR approach. After all, just like SR, IE suggests that theories should be statistically tested in a wide range of empirical domains. But insofar as process-tracing-based and interpretation-based mechanisms are concerned this assertion is clearly misguided: the entire purpose of process-tracing and interpretation is to capture *real* mechanisms in all their complexity; a goal that IE rejects.

¹⁹² Again, observable implications can also be non-correlational.

¹⁹³ For a similar approach, see George and Bennett (2005, 39). George and Bennett, however, do not discuss compatibility between interpretation and statistical analysis.

Insofar as rational choice-based mechanisms are concerned, the assertion remains faulty although the argument against it is somewhat more involved. Recall from the discussions above and in the previous chapter that rational choice models can be based on IE. IE involves specific approaches to rationality (employing the “as if” response), self-interest (employing a “thick objectivist” account), and methodological individualism (generating “rules of the game” solely from the interaction of individuals), which generally emphasize parsimony and generalizability over theoretical accuracy.

But given that rational choice models in political science are routinely developed *in order to make causal inference*, then IE can hardly serve as basis for epistemic unity in the discipline. Rational choice models that favour IE in regard to rationality, self-interest, or methodological individualism, and that *simultaneously* express a desire to make causal inference, are *philosophically incoherent*. Evidently, epistemic unity should not be based on philosophical inconsistency. Unless rational choice modelers are willing to eschew the goal of causation, only an SR-based approach to rationality (employing the “domain” response), self-interest (employing a “thin subjectivist” account), and methodological individualism (introducing “rules of the game” exogenously), provides philosophical consistency and provides a reasonable basis for breaching the epistemic divide between the nomothetic approaches of rational choice modeling and statistical analysis.

I return to the democratic peace literature in order to illustrate the potential for complementarity between statistical methods on the one hand, and process-tracing, rational choice modeling, and interpretation, on the other. I document where researchers

have demonstrated this type of complementarity and, where lacking, attempt to show how such complementarity can be achieved.

Process-Tracing, Statistical Methods, and the Democratic Peace

First, consider potential complementarity between process-tracing (which identifies chains of events and phenomena) and statistical methods. Elman (1997, 43) argues that process-tracing in democratic peace research is essential because, “the quantitative, statistical studies that prove that democracies virtually never go to war with one each other (the dyadic finding) – and are more peace-prone in general (the monadic finding) – remain controversial.” She argues further that:

Case studies can be used to develop, test, and refine theories...A small number of cases are investigated in detail to see if the theory’s causal explanations seem valid and if events unfold in the manner predicted. Many case studies involve tracing the decision making process to see if political actors speak and act in ways that are consistent with the theory’s logic...Case studies can identify antecedent conditions and particular circumstances under which the theory is likely to predict successfully...” (Elman 43)

Elman (36-42) suggests several ways that case studies can illuminate mechanisms. For instance, case studies can evaluate leadership orientation. Both normative and institutional theories hold that democratic and nondemocratic leaders act uniformly, but in fact democratic leaders may not have democratic values and autarkic leaders may be moderates. Case studies are required, argues Elman, to examine these nuances that are likely far from uncommon.

The SSR approach to the democratic peace supports Elman's primary claim: detailed and realistic explanations that reflect real-world processes are required to support the regularity between democracy and peace. However, Elman underestimates the power of quantitative analyses to test theoretical propositions and simultaneously overstates the ability of case studies to be the definitive tests of empirical regularities:

The quantitative empirical analyses that find that democracy is associated with peace are correlational studies, and provide *no* evidence that leaders actually consider the opponent's regime type in deciding between war and peace. These studies focus primarily on foreign policy outcomes and ignore the decision making process...In short, the challenge now facing democratic peace proponents and skeptics alike is to put the general theory to the test of detailed historical analysis: have leaders tended to act and think in ways consistent with the theory? (33, my emphasis)

Elman sees at best a minimal degree of complementarity between process-tracing and statistical analysis. The SSR approach, by contrast, suggests that theoretical refinements produced by process-tracing should be subjected to further testing against a range of empirical phenomena. By closely examining historical crises between and amongst democracies and non-democracies, many of the case studies contained in Elman's *Paths to Peace* generate useful refinements to normative and institutional theories, as well as scope conditions that suggest when and where democracy will engender peace. But these explanations for the democratic peace are never used to identify further hypotheses that are subject to statistical tests.

For instance, Elman (191-232) scrutinizes Finland's role in World War II, tracing the decision-making process that led democratic Finland to ally itself with autocratic Germany. Elman points out that Finland's role in World War II is held up by critics of the democratic peace – especially IR neorealists who emphasize power relationships – as an example of a democracy going to war with numerous fellow-democracies simultaneously, and therefore constitutes a strong disconfirmation of the democratic peace hypothesis. Democratic peace proponents retort that Finland never engaged the Allies militarily and rebuffed several requests from Germany. Elman's analysis of Finnish decision-making processes upholds a key element of the democratic peace hypothesis: that regime type, not neorealist calculus, determined Finland's strategic decisions.

But her analysis also reveals an important refinement: that the nature of the democratic process matters. More specifically, Elman concludes that:

Centralized democratic institutions will facilitate aggressive behavior when leaders prefer war. By privileging the executive over the legislative branch, such structures can lead the state down the road to war to war, perhaps even against fellow democracies. By contrast, decentralized democracies will be better able to constrain a leader bent on war, since the executive must share foreign policy making power with other groups, separating the foreign policy powers of the executive and the legislative branches will constrain war-prone leaders. (232)

Elman's theoretical refinement to democratic peace theory – one that is based on real decision-making processes – is important. But the conclusions that she draws from the single case are at once too strong and too broad. Aside from stating that "...the next round of [democratic peace] theory testing should identify its boundaries of applicability," Elman never suggests how her theoretical refinement should be tested against empirical reality. Clearly, Elman's analysis points to a particular scope condition – the extent to which a democracy is centralized – that should be accounted for in further statistical studies. What specific outcomes might we expect vis-à-vis the democratic peace if "degree of democratic centralization" makes a difference to decision-making processes? Are there threshold effects that one might expect vis-à-vis "degree of democratization?" Will it make a difference if two disputing democracies both have equally strong executive or legislative branches?

It may be unfair to suggest that Elman should have drawn out specific hypotheses around such questions and tested them statistically. These endeavours are likely beyond the scope of her particular piece of research. But by drawing strong conclusions on the basis of her single study, Elman appears to underestimate the importance of this approach. The SSR approach contends that neither statistical analyses nor process-tracing should alone form the basis for strong conclusions about causation. By integrating these two methods, however, conclusions about causation can become much stronger. The SSR approach, therefore, provides a basis for reconciliation to partially mend the historic rift between these two methods.

Interpretation, Statistical Methods, and the Democratic Peace

Like process-tracing, interpretation can also complement statistical methods. Recall from Chapter VII that interpretation generates findings about intersubjective meanings, which act as societal backdrops or contexts that enable regularities. Thus, Taylor (1994 [1971]) implicitly uses the “society of work” to explain how correlations between “integration,” “legitimacy,” and “social cohesion” are possible. And Price and Tannenwald (1996) explicitly use “prohibitory norms” to explain why regularities associated with deterrence theory are possible.

In a similar way, several interpretivists and constructivists use intersubjective meanings to explain regularities associated to the democratic peace. Hopf (1998, 192), for instance, suggests that interpretation, or what he prefers to call “conventional constructivism,” can be used to explain the democratic peace – what he calls, “an empirical regularity in search of a theory.”¹⁹⁴ Like most proponents of interpretation, Hopf contends that action is inherently meaningful. But unlike “critical constructivism” and many proponents of interpretation, Hopf (1991, 181-5) insists that “conventional constructivism” does *not* eschew the quest for causation.

Central to Hopf’s causal argument is the notion of *identity*. All states have identities – identities that they do not necessarily choose but rather are based on intersubjective understanding of who they are and what they stand for (Hopf 1998, 174-81). Thus,

¹⁹⁴ Constructing an interpretivist explanation of the democratic peace is not the primary intention of Hopf’s (1998) article.

“...during the Cold War, East European countries often understood the Soviet Union as Russian, despite the fact that the Soviet Union was trying hard not have that identity” (Hopf 1998, 175). Further, the identity of a state implies its preferences and consequent actions; and a state produces and reproduces its identity through these actions. For instance, “...U.S. appeasement in Vietnam was unimaginable because of the U.S. identity as a great power...Others observing the United States not only inferred U.S. identity from its actions in Vietnam, but also reproduced the intersubjective web of meaning about what precisely constituted that identity” (Hopf 1998, 172-3, 178).

Given that identities cause actions, Hopf (1998, 192) suggests that conventional constructivism is “perfectly suited” to providing an explanation of the democratic peace:

Its approach aims at apprehending how the social practices and norms of states construct the identities and interests of the same. Ergo, if democracies do not fight one another, then it must be because of the way they understand each other, their intersubjective accounts of each other, and the socio-international practices that accompany those accounts.

In short, Hopf suggests that the *identity* forms the core of an explanatory mechanism for the democratic peace.

Risse-Kappen (1995) also uses an identity-based constructivist argument to explain the democratic peace. Risse-Kappen (1995) contends that the dyadic proposition of the democratic peace is “under-theorized.” He argues for a social constructivist approach that refines the normative theory of the democratic peace. Normative theory, notes

Risse-Kappen (500-1), holds that people in a democracy share norms of “live-and-let-live,” and are unlikely to be led into conflict against persons in other democracies that share this norm. However, Risse-Kappen (501) argues that in a “dog-eat-dog world of anarchy...the perception of others as equally restrained might lead rational actors into adventurous foreign policies themselves or create the uncertainty that other rational leaders might come to the same conclusion.” Among other things, argues Risse-Kappen, normative theory, as it stands, cannot account for why a shared “live-and-let-live” norm does not generate more conflict than it prevents.¹⁹⁵

Risse-Kappen (502-9) employs a social constructivist argument to build on normative theory. His constructivist argument purports to explain, among other things, how “democracies ‘know’ that other democracies are equally peaceful and can, therefore, be trusted.” More specifically, Risse-Kappen contends that democratic norms constitute identities that are externalized by democracies, “...thus reinforcing the presumption of peacefulness.” It is the interaction between democracies that reinforces identities of peacefulness.

¹⁹⁵ Another potential problem that Risse-Kappen points to regarding normative theory has been largely disproved. Risse-Kappen (501) contends that if were true, as normative theory suggests, that “...aggressive behaviour is forced upon democracies by the mere existence of authoritarian regimes,” then “...one would expect to find clearer indications that militarized disputes between democracies and authoritarian systems are more often caused and initiated by the latter rather than the former.” He contends that “...there is not much in the empirical data to suggest this.” However, as noted above, Russett et al. (2001) and others have in fact since demonstrated that in mixed dyads, non-democracies are more often the initiators of conflict than democracies.

Allison (2001) provides a feminist constructivist explanation for the democratic peace. Arguing that “prevailing liberal explanations for [the] democratic peace, which focus primarily on democratic political institutions and culture, remain incomplete,” Allison suggests instead an ethics-based explanation centering on the “politics of care” (204). Allison argues that emotional ties that bind individuals are translated to the international realm: “The ethical foundation for peace understood as a forbearance from war as a means of conflict resolution lies not in the morality of international politics but rather in the moral choices of the individuals inhabiting the world’s nations” (207).

More specifically, Allison (211-4) contends that an ethic of “taking care of” and “caring about” permeates domestic politics of democracies. This is evidenced, for instance, in the vast number of programs aimed at taking care of the young, the poor, the homeless, etc. But Allison argues that domestic political goals ultimately depend on international cooperation: the politics of care depends therefore on peace. International disputes occur, but domestic ethical imperatives favour peaceful resolution: “it is the intention to reconcile inherent in the decision not to wage war that [is] significant because it renders international peace the almost certain outcome of deliberate choice in a relationship, rather than the potential result of an intractable evolutionary process” (206-7). In sum, the “politics of care” ensures that peace – not war – is the norm in international politics.

Other interpretivist arguments have also been advanced to explain the democratic peace (e.g., Cederman and Daase 2003, 13-22, Kahl 1999, Kivimäki 2001, and Williams 2001).

Largely absent from most of these studies, however, are suggestions as to how social constructivist explanations can be tested. There is little to no effort made at deriving testable hypotheses from these mechanisms in a manner consistent with SR. Given that these explanations are explicitly causal in nature, and advanced to explain regularities associated to the dyadic and/or monadic propositions of the democratic peace, it hardly seems like a stretch that these explanations could be tested against a range of other empirical phenomena, as advocated by the SSR approach.

Risse-Kappen (510-1) contends that measuring norms is “questionable” and therefore advises that only ethnographic or case-study research involving “detailed analyses of communication processes” is advisable in testing his constructivist theoretical mechanism. However, as argued in Chapter VII, the SSR approach does not treat norms as independent variables that are measured and correlated against dependent variables, but rather as enabling conditions that, if true, would “make possible” a range of empirical phenomena.

If Hopf (1998) and Risse-Kappen (1995) are right that democracies mutually constitute their identities as peaceful nations and therefore live together in peace, what else might we expect to find from these mutually constituted identities? Are there other empirical regularities or phenomena that we should expect? The theory of mutually constituted identities, for instance, suggests that in less modern times, when lines of international communication were not as open as they are today, peace between democracies would be

less likely.¹⁹⁶ In other words, one would expect to find that the correlation between democracy and peace varies with openness of international communication. One should ask the same types of questions of Allison's (2001) explanation regarding the "politics of care." If it is true that democratic nations rely on international support to reinforce their empathetic values, we might expect an increased likelihood of peace among those nations that have high domestic levels of "societal care" (e.g., strong social safety nets) and high levels of international interaction (e.g., frequent high-level political meetings). By demanding that these types of questions be asked and answered, the SSR approach provides the basis for epistemic unity between interpretation and statistical analysis in political science.

Rational Choice, Statistical Methods, and the Democratic Peace

Several democratic peace researchers have developed rational choice explanations of the democratic peace (e.g., Filson and Werner 2004, Schultz 2001).¹⁹⁷ As discussed above, Bueno de Mesquita et al. (1999) develop a game theoretic model of the democratic peace. The critical element of the model is that political leaders must satisfy a winning coalition to remain in power, a fact that determines how leaders fight wars and which wars they

¹⁹⁶ Risse-Kappen (1995, 510) notes that, according to Bachteler (1995), "Athens and Syracuse did not perceive each other as stable democracies and, thus, could not develop a collective identity preventing them from fighting each other."

¹⁹⁷ Filson and Werner (2004) are not concerned with the core monadic or dyadic propositions of the democratic peace, but rather with several empirical regularities associated with the democratic peace (e.g., that democracies are more likely to win the wars they initiate).

choose to fight. Democrats have to satisfy relatively large coalitions, whereas autocrats have to satisfy relatively small coalitions. Democracies are relatively more cautious about engaging in war than are autocracies, because democratic coalitions are more fragile than autocratic coalitions. However, for the same reasons, democracies are also likely to fight harder than autocracies once engaged in war. Regime type, in short, creates different incentives and strategies for leaders.

Bueno de Mesquita et al. (791) claim that their game theoretic model explains several regularities associated to the dyadic proposition of the democratic peace, including: (1) democracies are not unwilling to fight wars with nondemocracies¹⁹⁸; (2) democracies tend to win a disproportionate share of the wars they fight; (3) when disputes do emerge, democratic dyads choose more peaceful processes of dispute settlement than do other pairings of states; (4) democracies are more likely to initiate wars against autocracies than are autocracies against democracies¹⁹⁹; (5) in wars they initiate, democracies pay fewer costs in terms of human life and fight shorter wars than nondemocratic states; (6) transitional democracies appear more likely to fight than stable regimes; and (7) larger democracies seem more constrained to avoid war than do smaller democracies.

In addition, Bueno de Mesquita et al. (805) generate two “novel hypotheses” from their game theoretic model: (1) democracies try harder than autocracies once engaged in war;

¹⁹⁸ Clearly, Bueno de Mesquita et al. (1999) do not support the monadic proposition.

¹⁹⁹ This claim, attributed to Bennett and Stam (1998), supports their opposition to the monadic proposition but contradicts a good deal of evidence from supporters of the monadic proposition.

(2) democracies are not immune from wars of imperial expansion. They argue that existing empirical evidence supports both novel hypotheses. In a later article, Bueno de Mesquita et al. (2004) use their 1999 game theoretic model to refine these two hypotheses, positing that: (1) democratic leaders try harder to win wars than do autocrats, except under two conditions: (a) autocracies try hard when national survival is at risk, or (b) democracies do not try hard when the adversary is much weaker; (2) democracies show no reluctance to use force against other democracies in situations in which the rival democracy is expected to capitulate rather than fight back; (3) democracies show no reluctance to engage in wars against much weaker adversaries; (4) colonial and imperial expansion are especially attractive for democracies when the state's "democraticness" outpaces the size of its "winning coalition" (2-3). Bueno de Mesquita et al. (2004, 23) generate their own correlational tests of these refined hypotheses, concluding that "novel hypotheses derived from our explanation of the democratic peace are consistent with the observed evidence."

Taken together, Bueno de Mesquita et al.'s 1999 and 2004 articles on the democratic peace provide stellar examples of complementarity between the nomothetic approaches of rational choice modeling and statistical analysis: methods that are sometimes at odds with one another in the discipline of political science. His willingness to test his rational choice models against a wide range of empirical phenomena, especially novel empirical phenomena, is precisely what the SSR approach advocates. Given that one of the most penetrating criticisms of rational choice methodology made by Green and Shapiro (1994)

and Walt (1999) is that formal modelers routinely fail to test their theories against novel empirical evidence, Bueno de Mesquita's approach is especially relevant.²⁰⁰

However, as discussed in Chapter XII, Bueno de Mesquita's approach to rational choice modeling contains elements of both SR and IE. In order to make his rational choice models philosophically consistent with his goal of making causal inference, Bueno de Mesquita would have to resolve the inconsistencies in his approaches to rationality, self-interest, and methodological individualism, in favour of SR. Again, unless Bueno de Mesquita is willing to relinquish the goal of making causal inference, IE cannot provide a basis for epistemic unity between rational choice modeling and statistical analysis. The method of testing rational choice models against a wide range of empirical phenomena *with the intention of making causal inference* only makes sense in light of SR.

²⁰⁰ Green and Shapiro (1994) catalogue several "pathologies" that allegedly hamper rational choice modeling. A key pathology is that rational choice modellers typically engage in "post hoc theory development," or "curve fitting": rather than formulate bold predictions that are falsifiable by empirical evidence, rational choice modellers tend first to look at the empirical evidence, then design a rational choice model that fits that evidence. The SSR approach can help overcome this pathology as well as another pathology identified by Green and Shapiro: namely, "arbitrary domain restriction." Rather than post-hoc alterations that save a theory, the SSR approach suggests that a rational choice-based mechanism should "specify in advance particular reasons why a theory may not explain a certain class of phenomena" (Friedman 1995, 5). However, as argued in Chapter X, the SSR approach does not ascribe to Green and Shapiro's strict demarcation between theory testing and development: post-hoc reformulations have a place in the development of theories.

Two Further Elements of Epistemic Unity

The previous three sub-sections address a significant part of the first aforementioned disciplinary divide between ideographic and nomothetic methods (namely, the divide between process-tracing and interpretation on the one hand, and statistical methods on the other) as well as the second aforementioned disciplinary divide between proponents of nomothetic methods (namely, the divide between rational choice and statistical methods).

There remain two elements of epistemic unity requiring attention. First, in relation to the first disciplinary divide between ideographic and nomothetic methods, there is a divide between the case-study methods of process-tracing and interpretation on the one hand, and rational choice on the other. Second, there is the third disciplinary divide between proponents of ideographic methods; that is, between proponents of process-tracing and proponents of interpretation.

Chapter X alluded to the manner by which the SSR approach can help bridge the first of these two remaining divides. SR takes a unique approach to rationality, self-interest, and methodological individualism. What distinguishes SR from IE with respect to these three rational choice-related issues is the level of attention paid to theoretical accuracy. Unlike IE, SR-based rational choice modeling requires that rationality, self-interest, and methodological individualism be *contextualized*. SR rejects the “one size fits all” approach championed by IE in favour of modeling that reflects the *actual* nature of real-world politics. The in-depth empirical knowledge provided by process-tracing and

interpretation, in this sense, can be invaluable to formal modelers: process-tracing and interpretation can *provide* the requisite *context*. Ray's (1995) analysis of Bueno de Mesquita's rational choice modeling (see Chapter XII) makes clear that it is not beyond rational choice modelers to look to empirical case studies to develop or support their models. But again, to resolve the inconsistencies in his approaches to rationality, self-interest, and methodological individualism, further work on the democratic peace by Bueno de Mesquita would benefit from taking further advantage of process-tracing and interpretative case studies.²⁰¹

SR can also help bridge the gap between proponents of process-tracing and interpretation. The bulk of dissension between these two ideographic methods is attributable to differing approaches to the issue of causation. For some interpretivists, causation is a non-starter. They set themselves apart from proponents of process-tracing, arguing that they have fundamentally different goals. However, I have already argued that (as in the case of Taylor) causation is sometimes unavoidable; and, moreover, many interpretivists (as in the case of IR social constructivists) explicitly champion causation. To the extent that proponents of process-tracing and interpretation are united in their desire to make causal inference, there are many avenues for complementarity, as pointed out in Chapter VII. Given the importance of normative theory to the democratic peace, it is no small wonder

²⁰¹ In addition to using case study methods in conjunction with rational choice modeling, case study methods can also be used to test rational choice models. George and Bennett (2005, 9) point to Bates (1998) as an example of this type of cross-fertilization between methods.

that many of the case studies contained in Elman (1997) draw on insights from interpretative work on the democratic peace (e.g., see the chapter by Freedman, especially p.238). And it is not uncommon for interpretive works on the democratic peace to draw on process-tracing analyses (e.g., Risse-Kappen 1995, 508). This type of collaboration makes good sense in light of the SSR approach.

A General Point About Epistemic Unity

The above sections on the various methodological divides in political science raise an important general point about epistemic unity in the discipline. Without a common understanding of causation it becomes difficult to compare competing causal explanations. Gerring (2005, 166) makes this point nicely:

If causation means different things to different people then, by definition, causal arguments cannot meet. If *A* says that X_1 caused *Y* and *B* retorts that it was, in fact, X_2 or that *Y* is not a proper outcome for causal investigation, and they claim to be basing their arguments on different understandings of causation, then these perspectives cannot be resolved; they are incommensurable.

I contend that the SSR approach provides a unifying logic of causal inference. It provides researchers from distinct and traditionally competing branches of political science a common language by which to gauge causal arguments. The SSR approach makes causal arguments commensurable and therefore allows for progress in the discipline: “Insofar as we value cumulation in the social sciences, there is a strong *prima facie* case for a unified account of causation” (Gerring 2005, 166).

iii. Policy Prescriptions

In addition to helping generate better-trained practitioners and a measure of epistemic unity in political science, the SSR approach provides a third basis for progress in the discipline: its requirement of theoretical accuracy could help practitioners formulate better policy prescriptions. Like most social scientific disciplines, political science seeks to make itself useful to policymakers. This is not true of all political science nor should it be. Theory for theory sake is legitimate. But for much of political science, the ultimate goal is to steer policymakers to decisions that make the world a better, safer place (George and Bennett 2005, 7).

The democratic peace literature provides a good example. Democratic peace scholars routinely note that if the democratic peace is genuine, policymakers should be actively promoting democracy. And policymakers appear to be listening. Both the Clinton and Bush Administrations have made democratization a cornerstone of their foreign policies, in part because of the influence of the democratic peace literature.

But if the causal relationship between democracy and peace is genuine, policymakers should know more. Without an understanding of *how* democracy engenders peace, the specific policies around a foreign policy of democratization is, at worst, doomed to failure, and at best, unlikely to be as successful as it might otherwise be. What is it *specifically* about democracies that make them more peaceful than other regime types? What *exactly* are the mechanisms at work? Only SR provides a philosophical foundation

for answering these important questions. IE is interested in mechanisms solely insofar as they produce empirically adequate results, and HE is not at all interested in mechanisms in the first place. SR, conversely, attempts to identify mechanisms that accurately reflect real-world processes. More generally, as George and Bennett (2005, 7) note: “Highly general and abstract theories...which set aside intervening processes and focus on correlations between the ‘start’ and ‘finish’ of a phenomenon, are too general to make sharp theoretical predictions or to guide policy.”

SR recognizes that there are unlikely to be easy solutions to complex problems such as war. Groopman (2001) contends that for the past 30 years, America’s “war on cancer,” has mistakenly looked for the “silver bullet” cure without paying enough attention to the complexity of the underlying mechanisms that generate cancer. In a similar fashion, the SSR approach to war recognizes that simply enhancing democracy is unlikely to yield much success. Instead, policymakers should have access to the complex array of factors that will engender peace. In short, the SSR approach to studying the democratic peace, and political science in general, promises to yield more precise pragmatic advice to policymakers.

iv. Conclusions: The Big Payoff

In this chapter I have emphasized that the SSR approach to causation can advance the discipline of political science on several fronts. My approach promises to help yield better-trained practitioners, enhance epistemic unity, and generate more accurate information for policymakers. But it is important not to lose sight of the “big payoff”

that SR brings to the discipline. SR is the only philosophical foundation that is consistent with the desire of political scientists to make causal inference. Neither HE nor IE complements this goal.

One might argue that if philosophical unity were required, HE or IE would do just as well as SR. But unless political scientists decide to abandon their quest to make causal inference, this argument fails. Students and practitioners of political science are instructed to make causal inference but are often schooled concurrently in philosophical traditions that reject causation on epistemological and ontological grounds (or else, are sometimes not even introduced to these issue at all). Only SR holds that mechanisms are both relevant and literally true (or false), and hence only SR can serve as a legitimate philosophical foundation for making causal inference. And as this and the two previous chapters suggest, modifications that are required in political science to fully adopt the SSR approach are not as great as SRs typically claim. In short, my thesis suggests that relatively small adjustments to the discipline of political science could yield large gains.

Chapter XIV: Conclusions

The first section of this final chapter recaps my core arguments. The second section discuss how the SSR approach adds value to existing SR approaches in the social sciences. The third section anticipates and attempts to respond to some potential objections to my arguments. The fourth section considers some broad implications of my central claims. The fifth and final section offers some concluding remarks.

i. Summary

The desire to make causal inference, common to many political scientists, has been hampered by the enduring influence of empiricism. Causal relationships are sometimes treated as unexplained correlations (HE); or, more commonly, theoretical mechanisms invoked to explain correlations are relatively simplistic fictions without genuine explanatory power (IE). In short, political scientists typically espouse a black box approach to causation. To the extent that the discipline does embrace SR, it does so erratically and inconsistently. Moreover, methods that can identify accurate and detailed theoretical mechanisms are seen as separate, distinct, and often inferior from methods that generate correlational evidence, further impeding efforts at making causal inference. This thesis has developed a particular SR-based approach to causal inference that can help mitigate these problems and resolve the tension between the desire to make causal inference and the lack of a philosophical basis on which to do so. The SSR approach advocates thinking more deeply *inside* the black box.

More specifically, the SSR approach holds that the development and testing of accurate and realistic theoretical mechanisms that reflect real-world political processes is crucial to making causal inference. I contend that process-tracing, rational choice modeling, and interpretation, are methods that can identify three ontological properties – respectively, physicalness, agency, and intentionality – of social mechanisms, and generate findings that explain regularities in a unique fashion. Process-tracing identifies the ontological property of physicalness and generates findings that explain regularities as a chain of events or phenomena. Rational choice modeling identifies the ontological property of agency and generates findings that explain regularities as the aggregation or strategic interaction of individual choices. Interpretation identifies the ontological property of intentionality and generates findings that explain regularities as being “made possible” or “enabled” by intersubjective meanings.

But these methods provide only partial tests of the findings about the theoretical mechanisms that they generate. In order to make causal inference more robust, the SSR approach requires that the mechanisms be further tested in order to ensure that they operate as theorized and that they are literally true. The AfC, a core feature of scientific realism, provides the logic necessary to test in a robust manner the findings about social mechanisms identified by process-tracing, rational choice modeling, and interpretation. Thinking *inside* the black box entails embracing rather than eschewing correlational evidence.

The SSR approach to causal inference contrasts sharply with an approach based on HE. HE holds that theoretical mechanisms are neither relevant to scientific inquiry nor literally true or false. For HE, causation (to the extent that the term has any meaning) consists solely of unexplained correlations. The difference between the SSR approach, and HE, has important methodological implications for making causal inference. In contrast to HE, causal inference with my SSR approach is driven by theoretical mechanisms. Causal inference is made by deriving a range of observable implications from a theoretical mechanism and then testing these observable implications against the empirical record. Thus, unlike HE, researchers focus on *multiple* correlations (as well as other empirical phenomena).

The SSR approach to causal inference also contrasts with IE, although the contrast is much more nuanced. Theoretical mechanisms, according to IE, are relevant to scientific inquiry but have no truth value. IE champions parsimony and generalizability above all else, whereas the SSR approach is willing to sacrifice parsimony and generalizability on the altar of accuracy and reality. The difference between the SSR approach and IE over the ontological status of theoretical mechanisms has important methodological implications in relation to the value of process-tracing and interpretation, as well to the manner by which rational choice models are constructed. Unlike IE, the SSR approach treats process-tracing and interpretation as invaluable tools because they identify ontological properties of theoretical mechanisms that reflect the complexity of the real world. With regard to rational choice modeling, the SSR approach differs from IE in that

it treats assumptions relating to rationality, self-interest, and methodological individualism, as context-dependent.

I also assessed the “state of the art” of causal inference in political science and suggested how the SSR approach could help advance the discipline. To assess the discipline’s “methodological prescriptions” vis-à-vis causal inference, I closely examined DSI – widely touted as a leading methodological text in the discipline. To assess how causal inference is made in political science research, I examined the democratic peace – widely considered one of the discipline’s most mature research areas.

SRs in the social sciences – and, more specifically, political science – typically assert that social scientists take an HE-based approach to causal inference. My findings only marginally support this assertion. With regard to methodological prescriptions, I find that political scientists are taught an incoherent mix of HE, IE, and SR. I find that democratic peace researchers implicitly base their efforts at making causal inference in both IE and SR, and quite often an uneasy and inconsistent blend of the two. Compared to what existing SRs in the social sciences would seem to indicate, my analysis suggests that relatively minor adjustments are required on the part of political scientists to fully adopt the SSR approach to causal inference.

By adopting the SSR approach to causal inference, political science can evidently overcome the philosophical inconsistencies that characterize the discipline’s methodological prescriptions and its actual research. In addition, the SSR approach

promises to advance the discipline on at least three fronts. First, the SSR approach can yield better-trained practitioners by “making sense” of DSI. DSI offers sound research design advice, but the clarity of its advice is obstructed by its philosophical inconsistencies. Once these inconsistencies are resolved in favour of SR, researchers can better understand and even extend some of its techniques.

Second, the SSR approach can enhance epistemic unity by helping to mend several methodological divides in the discipline. The SSR approach suggests inherent compatibility between statistical analysis, process-tracing, rational choice modeling, and interpretation. This compatibility takes us beyond mere *triangulation* to genuine integration of methods.²⁰²

²⁰² On triangulation in the social sciences, see Bryman (1988, 123-33). On triangulation in political science, see Tarrow (1995, 473-4). At the other extreme, I do not necessarily contend, as does Gerring (2004), that case-study methods and quasi-experimental methods are substitutable – each with their own strengths and weaknesses, but both cut from the same cloth of rules regarding causal inference-making. Rather, I suggest that causal inference-making might require both case study methods (leaving aside for the moment, rational choice modeling) and quasi-experimental methods. (It is worth noting that Gerring (2004, 348) appear to believe, mistakenly in my estimation, that interpretivists such as Charles Taylor explicitly champion intentional analysis as an “integral part of causal analysis.”). Like George and Bennett (2005, 10, 35), I do not mean to suggest that every single researcher should be well-versed in, and employ, diverse methodologies in every single study. Rather, I want to encourage integration and collaboration between researchers and their respective specialties.

Third and finally, the SSR approach can advance the discipline by providing policymakers better advice. Because the SSR approach emphasizes the development of accurate and detailed theoretical mechanisms that reflect genuine political and social processes, it promises to give policymakers more precise information on which to act.

ii. Value Added to Existing Scientific Realist Approaches in the Social Sciences

My thesis builds on existing SR approaches, especially those relating to the social sciences. My thesis draws especially on works by Andrew Bennett, David Dessler, Jon Elster, Alexander George, and Daniel Little. The arguments I make develop this literature in several ways.

First, unlike existing approaches advocated by SRs in the social sciences, the SSR approach recognizes that mechanisms are not necessarily regularity-based. Little (1991, 14) defines his “causal mechanism thesis” as: “C is a cause of E [if] there is a series of events C_i leading from C to E, and the transition from each C_i to C_{i+1} is governed by one or more laws L_i .” Bennett and George (1997, 7) contend that each causal link in mechanism is inferred from a Hempelian covering law. With both these SR approaches, therefore, mechanisms are comprised of a series of regularities. The SSR approach includes the ontological property of mechanisms (physicalness) that explains regularities as a causal chain, but also recognizes two additional ontological properties of social mechanisms (agency and intentionality) that account for regularities without appeal to lower-level regularities. Rational choice modeling generates findings that explain regularities as the aggregation or interaction of individual decisions. Interpretation

generates findings that explain regularities as being “enabled” by intersubjective meanings.²⁰³

Second, when mechanisms are defined as a series of regularities, existing SR approaches in the social sciences fail to explain why mechanism is a relevant concept.²⁰⁴ I contend that even if a mechanism consists of a series of regularities, it is *treated* as a single mechanism in the logic of the AfC. That is, the discrete regularities are treated as a *single process* from which observable implications are derived and empirically tested.

Third, existing SR approaches fail to explain how mechanisms are tested beyond case-study analysis. I argue that although case studies can provide partial tests of mechanisms, further testing is required. I contend that the logic of the AfC provides this more robust

²⁰³ Little (1998) appears to accept rational choice-based explanations as mechanisms, and yet never reconciles this position with his 1991 definition of causal mechanisms. George and Bennett (2005) appear to recognize all three sources of social mechanisms, yet never distinguish them from one another in terms of *how* they each explain regularities.

²⁰⁴ Gerring (2005, 191), for instance, asserts that “*all* causal arguments strive for evidence of covariational (correlational) relationships between the putative X and Y *as well* as evidence of causal pathways between X and Y” (emphasis in original). But in the same breath, Gerring asserts that causal pathways “...are also covariational in nature...” because each link in the causal mechanism is established with correlational techniques. “Thus, to talk about mechanisms is also, necessarily, to talk about covariational patterns (‘correlations’)” (166). (Gerring (169) essentially describes himself as an SR, although he does not make SR philosophy a core component of his analysis.)

testing. Even though the AfC, in fact, might rely on correlational evidence, its logic gives equal weight to the importance of mechanism.

Fourth, unlike some existing SR approaches that treat mechanisms as an end to explanation, I contend that mechanisms are continuously in want of deeper explanation. Contra these approaches, I argue that causation is given meaning by epistemic communities: causal status is not conferred on a relationship by discovering a fluid process between cause and effect; rather, causal status is conferred on a relationship when a community of scholars widely agree that a mechanism has surpassed a threshold – what I have called a confidence level in causation (CLIC) – beyond which further explanation is not required (even though it still might be desirable).

My argument about epistemic thresholds further serves to resolve an outstanding puzzle that plagues existing SR approaches: If mechanisms are comprised of regularities, and regularities are explained by mechanisms – as Little (1991) and Bennett and George (1997) argue – then the relationship between the two has no end: there is an infinite regress of mechanisms and regularities. But then on what basis do these SRs contend that causation is a meaningful term? My notion of epistemic thresholds, or CLICs, suggests that even if the relationship between mechanisms and regularities is theoretically boundless, there are pragmatic boundaries erected by scholarly consensus.

Fifth, some existing SR approaches fail to recognize the importance of ontology: that is, they fail to recognize that unless mechanisms are real, they cannot cause. The SSR

approach recognizes that the social world is comprised overwhelmingly of *ideas*, which are, by definition, not mind-independent, as SR would seem to require. I provide philosophical reasons for how ideas can be considered literally true or false. I also show how the logic of the AfC can establish whether or not unobservables are literally true.²⁰⁵

Because existing SR approaches sometimes fail to acknowledge the importance of ontology, they are sometimes indistinguishable from non-SR approaches that also argue for the importance of mechanism. Many social scientists contend that regularities must be supplemented by mechanisms without necessarily adopting a SR approach (e.g., see Shaffer 2002 or Sambanis 2004). What sets the SSR approach apart is that I provide reasons for why the ontological status of mechanism should be central to a SSR approach (because what does not exist, cannot cause) and provide philosophical reasons and methodological guidelines for establishing the literal truth or falsehood of unobservables.

Perhaps most importantly, my treatment of ontological status refines the debate over causation in important ways. Whereas most SRs define the debate largely as one

²⁰⁵ Some SRs in the social sciences do recognize the importance of ontology. George and Bennett (2005), for instance, explicitly discuss the ontology of unobservables. However, I believe that George and Bennett concede too much on this front to IEs. Their discussion on p.143, for instance, appears to suggest that ultimately we are stuck with IE: “At the frontier of our knowledge at any given time, our theoretical commitment to molecules, atoms, quantum mechanics, or string theory resemble an ‘as if’ assumption about the underlying mechanism at the next level down.” This statement and the accompanying passage appear to relinquish the belief, held by SRs, that unobservables have a truth value.

between HE and SR, my analysis of the role that ontology plays in debates over causation leads to the inclusion of IE: and as argued, much of the tension in political science is actually between IE and SR.

Sixth and finally, existing SR approaches fail to distinguish between the causal mechanism and causal power approaches. The two are often conflated in the literature and some authors seem to adopt both without recognizing the distinction. This thesis attempts to make the distinction clear and argues that the causal mechanism is superior to the causal power approach because of significant philosophical flaws in the latter.

iii. Potential Objections

In this section I attempt to anticipate and respond to several potential objections to my arguments.

Can Correlations not Surpass Epistemic Thresholds?

Contra my argument that mechanisms are required to explain regularities, some historical examples from the natural sciences would seem to suggest otherwise. For instance, the causal nature of the relationship between smoking and cancer would seem to have been established many years before mechanisms were discovered to explain this correlation. Similarly, the mechanism that explains the relationship between aspirin and headache relief was discovered many years after causal status would appear to have been conferred on this correlation. If epistemic thresholds exist for mechanisms, do they not exist for regularities? If statistical tests continue to uphold a correlation despite there being no

compelling theoretical mechanism, will a scientific community not, at some point, come to a general consensus that the correlation is causal?

In fact, in actual scientific practice, it may *not* be common for unexplained correlations to pass beyond epistemic thresholds. As Thagard (2000) notes (see Chapter III), the scientific community rejected the bacterial theory of ulcers until a compelling mechanism was found to explain strong correlations between the presence of the bacterium, *Helicobacter Pylori* and ulcers. Similarly, the tobacco industry only capitulated to legal pressures once the mechanism explaining the correlation between smoking and cancer was established.²⁰⁶ For years prior, the industry successfully avoided litigation by arguing that “correlation does not entail causation.”²⁰⁷

Nevertheless, one might still reasonably argue that my position sets a double standard: unexplained correlations should have epistemic thresholds just as mechanisms do. I am prepared to acknowledge that in some cases – if correlational evidence is strong and persistent, and if all *conceivable* confounding variables have been controlled – mechanisms may not be required. But this concession requires an important caveat.

²⁰⁶ The “tobacco” case differs from the “ulcer” case in that the latter contained a good reason (namely, stomach acid) to doubt the causal connection between bacteria and ulcers. No such reason existed to doubt the correlation between smoking and cancer. Nevertheless, in both cases an epistemic threshold was only passed once a mechanism was supplied.

²⁰⁷ This point is based on my own observations about the timing of the tobacco industry collapse. Further research would be required to firmly establish the linkages that I draw.

Unlike many natural scientists, social scientists usually cannot conduct laboratory or randomized experiments. Thus, mechanisms might be more important in the social than in the natural sciences. The difference between controlled experiments and quasi-experiments is one of degree, not kind. Philosophers recognize that even with laboratory and randomized experiments, it is theoretically impossible to control for *all* potentially confounding factors.²⁰⁸ The ability of quantitative methods to determine causation is best seen as a continuum, beginning on the left with simple bivariate methods, through multivariate methods, through the causal models associated with the Artificial Intelligence community at Carnegie Mellon University (see McKim and Turner 1997), through to randomized and laboratory experiments on the right of the continuum. As one moves from right to left on this continuum, mechanisms become increasingly important in determining causation.²⁰⁹

²⁰⁸ We might *conceive* of many potentially confounding factors, but we can never be certain that *all* confounding factors have been controlled. This is true even for laboratory and randomized experiments. Thus Koslowski (1996, 12) argues: “[T]he Humean principle according to which one can identify the cause of an event by examining what covaries with it is an important principle of causal reasoning. However, in any particular case, given all the possible covariations in the world, one must rely on a non-Humean index, knowledge of mechanism, to avoid being overwhelmed by the irrelevant ‘noise,’ that is, to distinguish causal correlations from those that are merely artifactual.” (Moreover, as noted in Chapter I, some complexity theorists argue that control becomes virtually impossible when outcomes are highly sensitive to initial conditions.)

²⁰⁹ According to McKim (1997, 16), Clark Glymour – one of the key scholars from the Carnegie Mellon group – contends “that there are actually many situations in which the statistical methods that he champions will yield information about causal structure superior to that which could, in

Do Failed Experiments Mean Faulty Mechanisms rather than Faulty Regularities?

I have argued that to test a regularity, we identify a theoretical mechanism, derive observable implications, and test these implications against the empirical evidence. Further, I have argued that if we move horizontally to a new experiment that does not affirm the theoretical mechanism, the causal status of the original regularity is called into question. However, is it not possible that the failure to affirm a theoretical mechanism against additional observable phenomena is an indication that the theoretical mechanism, *not* the original regularity, is questionable?

In fact, this *is* possible, and probably fairly common. The point remains, however, that we should only rarely rest content with unexplained regularities. As the scientific episode surrounding the bacterial theory of ulcers suggests, a plausible mechanism is usually required to affirm the causal status of correlational evidence. Thus, disconfirming empirical evidence may call both a given regularity *and* its proposed theoretical mechanism into question: casting the theoretical mechanism aside hardly entails accepting the regularity.

If there are no “Necessary Connections,” how does SR Differ from HR?

My argument against an end of explanation would seem to suggest that there are no necessary connections between unobservables, and hence my position on causation

principle, be attained by experimental methods.” Similarly, Leamer (1983) contends that correlation-based methods can be as effective as experimental methods in ascertaining causation.

appears indistinguishable from that of HR – that is, my argument should force me to be skeptical about the nature of causation.

This potential objection requires two responses. First, as argued, I contend that rational choice-based and interpretation-based mechanisms are not regularity-based, so my position differs fundamentally from that of HR. Second, my position on epistemic thresholds suggests that necessary connections are not required to establish causation. Even if, at bottom, theoretical mechanisms are comprised of unobservable entities, what is crucial to establishing causation from my perspective is that enough evidence for the mechanism is borne out in the empirical evidence to allow for an epistemic consensus.

Does Path Analysis not Identify Mechanisms?

Path analysis is a statistical method that attempts to establish the causal chain between an IV and a DV. As such, this method might appear to generate findings about chains of events and phenomena with purely correlational evidence, thereby vitiating my argument for the importance of mechanism. As Gerring (2004, 349) puts it, “There is a general perception – common at least among protagonists of the case study – that cross-unit studies entail large black boxes with no peepholes and that, therefore, they *must* be supplemented by the in-depth analysis of key units. This is not always the case.”²¹⁰

However, I contend that correlational evidence must always be tested by the logic of the AfC. Thus, each causal linkage in a path analysis must be tested according to the

²¹⁰ Gerring’s (2004, 349) own position is that “case studies are more likely to shed light on causal mechanisms and less likely to shed light on causal effects” than are quasi-experimental methods.

principles of the SSR approach: that is, a theoretical mechanism for each correlational step must be established. It might also be possible to treat the intervening causal linkages in a path analysis (that is, the correlational links between the key IV and key DV) as a theoretical mechanism in the very same way that I have argued is done for process tracing: observable implications of the mechanism in its entirety are derived and tested against a range of empirical domains.

How can the SSR Approach Choose Between Competing Theoretical Mechanisms?

The SSR approach emphasizes the development of accurate and detailed mechanisms that can limit generalizability. Yet I have also argued that theoretical mechanisms are tested in large part by their ability to explain phenomena in a range of empirical domains. This suggests that there is no clear way to choose between competing theoretical mechanisms. A researcher will have to continuously balance the need for accuracy against the requirement of generalizability: there are no specific criteria for choosing between a relatively detailed mechanism that is relatively less portable and a relatively less detailed mechanism that explains relatively more. IE, on the other hand, appears to have a clear criterion for choosing between competing theoretical mechanisms: the one that helps predict the most empirical phenomena wins out.

This potential objection, however, is vitiated by problems with *verificationism*. It is philosophically unsound to attempt to confirm a theoretical mechanism by looking solely for affirming evidence, because there are an infinite number of observable implications consistent with any theoretical mechanism. More to the point, for every theoretical

mechanism there is an infinite number of “easy tests” that will affirm its validity. Therefore it is not philosophically sound to simply compare theoretical mechanisms based on the sheer number of observable implications that they explain: a theoretical mechanism that “passes” twenty “easy tests” should not be favoured over one that “passes” five “hard tests.” This means that the explanatory power of a theoretical mechanism is not measured by simply counting the number of observable implications that it can explain. Rather, the number of observable implications that a theoretical mechanism can explain must be balanced against a qualitative assessment of the difficulty of these “tests.” In this sense, the SSR approach to testing is not unique: epistemic communities usually have a role to play in choosing between competing theoretical mechanisms.

Further, the criteria that IE employs to choose between theoretical mechanisms also include parsimony: theories that explain more with less are preferred to those that explain less with more. But determining whether one theoretical mechanism is more parsimonious than another is hardly a perfectly objective task. Here too, epistemic consensus is usually required. Moreover, parsimony is only of several criteria by which social scientists adjudicate competing theories; and because there are no hard rules by which one can prioritize these criteria, epistemic consensus plays a role here too. Thus again, the SSR approach may not be unique in its reliance on epistemic thresholds.

Can Case Studies Produce Generalizable Findings?

It might be objected that process-tracing and interpretation are unlikely to identify mechanisms that are portable beyond the very cases in which they are used: that the *sine qua non* of these methods is to unearth processes that are historically and culturally *unique*. This objection has merit, but case study research is not necessarily confined to a single place, time, or event. In fact, case study researchers routinely look for patterns across numerous individual studies (e.g., see Homer-Dixon 1999 or Sambanis 2004). This approach to case study research entails separating out systematic from nonsystematic factors: an arduous task, but one that is achievable. Moreover, case study research is routinely conducted on issues that cut across time and space. Consider, for instance, the constructivist literature on the democratic peace, discussed in Chapter XII. In short, case studies – whether single or multiple in number – can generate knowledge that is applicable to universal or general phenomena. Indeed, Gerring (2004, 342) defines a case study in precisely these terms: “...I propose to define the case study as *an intensive study of a single unit for the purpose of understanding a larger class of (similar) units*” (emphasis in original).²¹¹

Should we not Eschew Models that are just as Complex as Reality?

²¹¹ George and Bennett (2005, 32, 109-115, 233-262) contend that case studies can generate universal generalizations, although they are far more likely to generate contingent generalizations that are highly circumscribed. I do not necessarily disagree with them on this point, as even physics has difficulties sustaining truly universal generalizations (see Cartwright 1983).

Philosophers of science generally agree that we want to avoid theoretical mechanisms that are as complex as the reality that they are intended to model (Tetlock and Belkin 1996). It might be argued that my focus on detailed and accurate mechanisms contradicts this principle. I maintain that detail is important, but researchers must sort out *systematic* from *nonsystematic* factors. Background noises that alter relationships in random ways across time and space should not be included as part of theoretical mechanisms. I acknowledge, however, that in practice, sorting out systematic from nonsystematic factors can be very difficult.

It may also be objected that as theoretical mechanisms become increasingly complex, testing them becomes pragmatically more difficult. The complexity of a mechanism can so overwhelm a data set as to make testing pragmatically impossible. As Forbes (2004, 66) notes, "...despite faster computers, larger data archives, and more powerful statistical methods, it remains true that realistic causal models of political processes often far outrun our ability to test them in any rigorous way." I do not deny this practical limitation to my emphasis on complexity, and researchers should always do their best to ensure that data sets can withstand the strains imposed by complex models. But to *insist* that researchers tailor theoretical mechanism to match the robustness of data sets is to have the tail wag the dog: it is an irresponsible and methodologically unsound practice. (Forbes does not suggest this practice.)

Is My Tri-Partite Definition of Mechanisms All-Inclusive?

It might be argued that my tripartite definition of mechanisms is not all-inclusive. Can social mechanisms consist of something more than physicalness, agency, and intentionality? I would argue that this categorization is, in fact, all-inclusive. However, the ontological properties of social mechanisms should not be confused with the methods used to identify these properties. For instance, I certainly acknowledge that process-tracing is not the only method that can identify physicalness and generate findings that explain correlations as a causal chain. Another social scientific method that can generate similar findings is Michael Scriven's *Modus Operandi Method* whereby, a mechanism "...is an associated configuration of events, process, or properties, usually in time sequence, which can often be described as the 'characteristic causal chain' (or certain distinctive features of the chain) connection the cause with the effect" (Scriven 1976, 105).²¹² Similarly, rational choice modeling and interpretation may not be the only methods available for identifying, respectively, agency and intentionality.²¹³

²¹² Scriven's *Modus Operandi Method* resembles the scientific method of *biomarking*, whereby a plausible cause is 'marked' with a biological indicator (e.g., a dye) and followed through time and space to see if it manifests itself in the alleged effect. For more details on biomarking, see Vineis and Porta (1996) and Specter (2001).

²¹³ It is also possible that theories of agency are not fully represented by rational choice modeling (which is, after all, a theoretically-informed method). Thus, as noted in Chapter VI, Bennett (2003, 12) and Elster (1998) identify additional theories of agency that are important to the development of mechanisms.

Does the SSR Approach Really Differ from that Advocated by DSI?

Mainstream political science methodologists might complain that I overstate their averseness to using process-tracing and interpretation to generate testable causal mechanisms. After all, DSI itself provides examples of how process-tracing and interpretation can be used to generate testable hypotheses. There are critical differences, however, between my approach to generating testable mechanisms and that championed by DSI. DSI's approach is to turn every insight from case study research into a statistical test. Thus, DSI (39-40) asserts that Clifford Geertz's insight into the distinction between winking and twitching (the difference depends on social context) "is best expressed as a causal hypothesis...the hypothetical causal effect of the wink on the other political actor is the other actor's response given the eyelid contraction minus his response if there were no movement (and no other changes)." Similarly, with regard to process-tracing, DSI (85-6) maintains that although "an emphasis on causal mechanisms makes intuitive sense," demonstrating "the causal status of each potential linkage in such a posited mechanism, the investigator would have to define and then estimate the causal effect underlying it." Hence, causal effect is "logically prior" to causal mechanism.

The SSR approach advocates a much more holistic approach to process-tracing and interpretation research. Rather than parse every insight into a discrete statistical relationship, my approach asks, "What can be learned from a case study and what are its implications for other empirical phenomena?" DSI and mainstream methodologists in political science want to make statisticians out of qualitative researchers. The SSR approach takes case study research at face value. DSI advocates unity between

qualitative and quantitative research on the basis that if case study researchers put their minds to it, they can approximate the logic of their would-be quantitative brethren.

Conversely, I suggest that case study research in and of itself has value: value that is, to a much greater extent than DSI would allow, independent of the correlative approach favoured by mainstream political science.

Must Unobservables be Literally True, if Prediction is the Goal of Policymaking?

If IE can yield accurate predictions, it might reasonably be asked why anything else should matter to policymakers. I have two responses to this potential objection. First, I am skeptical that IE ultimately *can* generate more accurate predictions than SR; it stands to reason that as our knowledge of mechanisms increase, predictions will become more accurate, if less generalizable. Second, even if IE can yield accurate predictions, there may be no avenue for policy intervention if the mechanisms are merely useful fictions. IEs are happy to rest content with getting the right answer for the wrong reason. SR at least is dedicated to identifying real processes that should be amenable to policy intervention.

Sambanis (2004, 273) makes both these points in relation to his case study research on civil war. He notes that past quantitative studies on the onset of civil war “make the right predictions, but for the wrong reasons.” “If large-N studies make incorrect assumptions about causal paths,” notes Sambanis, “they will lack explanatory power. Moreover, the inferences drawn from these models cannot yet inform policy except in a very general and indirect way.”

Does my Analysis of DSI not presuppose the need for SR?

In Chapter XIII, I suggest several ways by which SR can make sense of, and even extend, some of DSI's pragmatic advice for increasing the N of a small-N study. But do my suggestions not *presuppose* the need for SR? That is, if one accepts IE, is it not possible that DSI's pragmatic advice does make perfect sense and does not require any extensions? The problem with this potential objection is that if one accepts IE as a philosophical foundation on which to build research techniques, then one cannot reasonably accept DSI's explicit rejection of "noncausal explanation." My amendments to DSI's research design are consistent with DSI's explicit endorsement of genuine causal explanation; to accept DSI's research design without these amendments creates philosophical inconsistencies.

iv. Broad Implications

The SSR approach could be extended beyond its application here in at least four different ways, although the first three of these potential implications require much greater research.

Extension to Other Social Science Disciplines

I have hinted that the findings of my thesis might have implications that extend beyond political science to the social sciences more generally. At least two conditions are required for this extension to other social science disciplines. First, the social science disciplines should contain researchers who work with both quantitative and qualitative

methods: that is, some combination of statistical analysis, process-tracing, rational choice modeling, and interpretation. Second, like political science, the opportunity to use randomized experiments should be limited. Any social science that meets these conditions might be able to benefit from the SSR approach to making causal inference.

Collaboration between Political Science and Other Social Science Disciplines

The SSR approach provides a basis for a certain measure of epistemic unity within the discipline of political science, but it also suggests the possibility of greater unity *between* political science and other disciplines. My emphasis on precise and detailed theoretical mechanisms suggests that political scientists should consider increasing their collaboration with their brethren in fields such as sociology, anthropology, and economics. Political processes are obviously intertwined with other social processes. Researchers from all social science disciplines can potentially contribute to our knowledge of political processes.

Causal Weighting

The SSR approach might also have implications for other aspects of causation, particularly the issue of *causal weighting*.²¹⁴ It is generally agreed that statistical methods can assess the relative importance of variables in determining an outcome. The coefficients in a regression analysis, for instance, provide numerical indications of the relative importance of a regression equation's IVs. Some scholars (e.g., Homer-Dixon

²¹⁴ For discussions on causal weighting, see Frey (1976), Hammond (1977), Lewontin (1976), Martin (1972), Pork (1985), and Sober (1988).

2000, 279-309), however, contend that the ability of statistical analysis to provide accurate causal weights is severely restricted by interaction effects, nonlinearities, and threshold effects. They claim that statistical methods such as regression analysis are ill equipped to handle relationships between variables that are not additive and linear.

The SSR approach might be useful because theoretical mechanism could help determine the *nature* of causal relationships: that is, whether they are additive or interactive, or whether they are linear or nonlinear.²¹⁵ Knowing *why* an IV correlates to a DV should help determine what the functional relationship between these variables looks like, and what other factors might interact with the IV to produce the DV.²¹⁶ For instance, if, as institutional theory suggests, democracy and peace correlate because democratic political structures impose high costs on leaders that bring their country to war unnecessarily, then a more detailed understanding of the nature of these political structures might help us determine what factors might override, magnify, or lessen the impact of political institutions (possible interaction effects), how differences in democratic institutions generate a varying potential to exact costs on political leaders (possible nonlinearities), and at what point different types of democratic institutions begin and stop having potential to exact these costs (possible threshold effects).

²¹⁵ Sambanis (2004, 260) contends that case studies can help determine interaction effects, as do George and Bennett (2005). But this argument differs from the one that I make here.

²¹⁶ For an example of using process-tracing to determine causal weight, see Runkle (2003, 72, 285-304).

Bringing Philosophy Back In

Practitioners of political science sometimes studiously avoid the epistemological and ontological underpinnings of their research methods because they erroneously believe them to be of little import. As George and Bennett (2005, 127) note: "...[P]racticing social scientists can be too disengaged from developments in the philosophy of science. Many scholars in the field of international relations, for example, appear to have become too removed from these developments." MacDonald (2003) makes a similar case. If nothing else, my hope is that this thesis has highlighted the importance of connecting philosophy to practice in the discipline of political science. This is an issue that extends far beyond the narrow slice of causation that I have attempted to deal with in this thesis.

v. Concluding Remarks

Much of political science is devoted to making causal inference, but powerful philosophical currents – most notable, HE and IE – that take a black box approach to causation often thwart the effort. If the discipline of political science is to advance in its quest to make causal inference, it requires a philosophical underpinning that complements this goal. The SSR approach to causal inference provides such a philosophical approach: sometimes, new approaches require that we think more seriously *inside* the box.

References

Ahn, Woo-Kyoung, Charles W. Kalish, Douglas L. Medin, and Susan A. Gelman. 1995.

“The Role of Covariation Versus Mechanism Information in Causal Attribution.”

Cognition 54: 299-352.

Allison, Juliann Emmons. 2001. “Peace Among Friends: A Feminist Interpretation of the

‘Democratic Peace’.” *Peace & Change* 26: 204-22.

Almond, Gabriel A. 2004. “Who Lost the Chicago School of Political Science?” 2004.

Perspectives on Politics 2: 91-3.

Armstrong, D.M. 1968. *A Materialist Theory of the Mind*. London: Routledge.

Aronson, Jerrold. 1971. “On the Grammar of Cause.” *Synthese* 22: 414-30.

Aronson, Jerrold. 1982. “Untangling Epistemology from Ontology in Causation.”

Erkenntnis 18: 293-305.

Babbie, Earl. 1998. *The Practice of Social Research* (Eighth Edition). Belmont, CA:

Wadsworth Publishing Company.

Barkawi, Tarak, and Mark Laffey, eds. 2001. *Democracy, Liberalism, and War: Rethinking the Democratic Peace Debate*. Boulder, CO: Lynne Rienner Publishers.

Bates, Robert, Avner Greif, Margaret Levi, Jean-Laurent Rosenthal, and Barry Weingast. 1998. *Analytical Narratives*. Princeton, NJ: Princeton University Press.

Bennett, Andrew. 1997. "Lost in the Translation: Big (N) Misrepresentations of Case Study Research", Thesis presented at the 38th Annual Convention of the International Studies Association in Toronto, March 18-22, 1997.

Bennett, Andrew. 2003. "The Mother of All 'Isms:' Organizing Political Science Around Causal Mechanisms." Thesis presented at the International Studies Association Conference, Portland, Oregon, March 1, 2003.

Bennett, Andrew, and Alexander George. 1997. "Process Tracing in Case Study Research." Thesis presented at the MacArthur Foundation Workshop on Case Study Methods, Belfer Center for Science and International Affairs (BCSIA), Harvard University, October 17-19, 1997.

Bennett, Scott D., and Alan C. Stam III. 1998. "The Declining Advantages of Democracy: A Combined Model of War Outcomes and Duration." *Journal of Conflict Resolution* 42: 344-66.

Bennett, Stephen Earl. 2002. "'Perestroika' Lost: Why the Latest 'Reform' Movement in Political Science Should Fail." *PS: Political Science & Politics* 35: 177-9.

Benoit, Kenneth. 1996. "Democracies Really Are More Pacific (in General): Reexamining Regime Type and War Involvement." *The Journal of Conflict Resolution* 40: 636-57.

Bhaskar, Roy. 1978. *A Realist Theory of Science*. New York: Harvester Wheatsheaf.

Bhaskar, Roy. 1979. *The Possibility of Naturalism*. London: Routledge.

Brady, Henry E. and David Collier, eds. 2004. *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. Lanham: Rowman & Littlefield Publishers, Inc.

Bremmer, Stuart A. 1992. "Dangerous Dyads: Conditions Affecting the Likelihood of Interstate War, 1816-1965." *Journal of Conflict Resolution* 36: 309-41.

Brown, James Robert. 2001. *Who Rules in Science? An Opinionated Guide to the Wars*. Cambridge, MA: Harvard University Press.

Bryman, Alan. 1988. *Quantity and Quality in Social Research*. London: Unwin Hyman.

Bueno de Mesquita, Bruce. 2002. "Domestic Politics and International Relations."

International Studies Quarterly 46: 1-9.

Bueno de Mesquita, Bruce and David Lalman. 1992. *War and Reason*. New Haven, CT:

Yale University Press.

Bueno de Mesquita, Bruce, James D. Morrow, Randolph M. Siverson, and Alastair

Smith. 1999. "An Institutional Explanation of the Democratic Peace." *American Political*

Science Review 93: 791-807.

Bueno de Mesquita, Bruce, James D. Morrow, Randolph M. Siverson, and Alastair

Smith. 2004. "Testing Novel Implications from the Selectorate Theory of War." *World*

Politics 56: 363-88.

Cartwright, Nancy. 1989. *Nature's Capacities and Their Measurement*. Oxford: Oxford

University Press.

Cartwright, Nancy. 1983. *How the Laws of Physics Lie*. Oxford: Oxford University Press.

Cederman, Lars-Erik. 2003. "Modeling the Size of Wars: From Billiard Balls to

Sandpiles." *American Political Science Review* 97: 135-50.

Cederman, Lars-Erik, and Christopher Daase. 2003. "Endogenizing Corporate Identities: The Next Step in Constructivist IR Theory." *European Journal of International Relations* 9: 5-35.

Checkel, Jeffrey T., "The Constructivist Turn in International Relations Theory." *World Politics* 50: 324-48.

Chernoff, Fred. 2002. "Scientific realism as a meta-theory of international politics." *International Studies Quarterly* 46: 189-207.

Collingwood, R.G. 1940. *An Essay on Metaphysics*. Oxford: Clarendon Press.

Collingwood, R.G. 1946. *The Idea of History*. Oxford: Oxford University Press.

Coleman, William D., and Melissa Gabler. 2002. "Agriculture Biotechnology and Regime Formation: A Constructivist Assessment of the Prospects." *International Studies Quarterly* 46: 481-506.

Copeland, Dale C. 1999. "The Constructivist Challenge to Structural Realism." *International Security* 25: 187-212.

Crease, Robert R. and Charles C, Mann. 1986. *Second Creation*. New York: Macmillan Publishing.

Dahl, Robert A. 1961. "The Behavioral Approach in Political Science: Epitaph for a Monument to a Successful Protest." *American Political Science Review* 55: 763-72.

Davidson, Donald. 2001 [1963]. "Actions, Reasons, and Causes." *Essays on Actions and Events*. Oxford: Oxford University Press.

Davis, Kingsley. 1959. "The Myth of Functional Analysis as a Special Method in Sociology and Anthropology." *American Sociological Review* 24: 752-72.

Dessler, David. 1991. "Beyond Correlations: Toward a Causal Theory of War." *International Studies Quarterly* 35: 337-55.

Dessler, David. 1999. "Constructivism Within a Positivist Social Sciences." *Review of International Studies* 25:123-137.

Diesing, Paul. 1971. *Patterns of Discovery in the Social Sciences*. Chicago: Aldine-Atherton.

Dinsmoor, James A. 1999. "Foreword" In *The Philosophical Legacy of Behaviorism*, ed. Bruce A. Thyer. Dordrecht: Kluwer Academic Publishers.

Dion, Douglas. 2003. "Evidence and Inference in the Comparative Case Study." In *Necessary Conditions: Theory, Methodology, and Applications*, Gary Goertz and Harvey Starr, eds. Lanham: Rowman & Littlefield Publishers, Inc.

Dixon, William J. 1994. "Democracy and the Peaceful Settlement of International Conflict." *American Political Science Review* 88: 1-17.

Dowe, Phil. 1992. "Wesley Salmon's Process Theory of Causality and the Conserved Quantity Theory." *Philosophy of Science* 59: 195-216.

Dowe, Phil. 2000. *Physical Causation*. Cambridge: Cambridge University Press.

Dray, William. 1957. *Laws and Explanations in History*. Oxford: Oxford University Press.

Dray, William. 1989. *On History and Philosophers of History*. Leiden: E.J. Brill.

Eckstein, Harry. 1975. "Case Study and Theory in Political Science." In *Strategies of Inquiry*, eds. Fred I Groenstein and Nelson W. Polsby. Reading, MA: Addison-Wesley Publishing Co.

Earman, John. 1986. *A Primer on Determinism*. Dordrecht: D. Reidel Publishing Company.

Elman, Miriam Fendius, ed. 1997. *PatHE to Peace: Is Democracy the Answer?*

Cambridge, MA: The MIT Press.

Elman, Miriam Fendius. 1997. "Finland in World War II: Alliances, Small States, and the Democratic Peace." In *Paths to Peace: Is Democracy the Answer?*, ed. Miriam Fendius

Elman. Cambridge, MA: The MIT Press: 191-232.

Elster, Jon. 1983. *Explaining Technical Change: A Case Study in the Philosophy of Science*. Cambridge: Cambridge University Press.

Elster, Jon. 1989. *Nuts and Bolts for the Social Sciences*. Cambridge: Cambridge University Press.

Elster, Jon. 1998. "A plea for mechanisms." In *Social Mechanisms: An Analytical Approach to Social Theory*, eds. Peter Hedström and Richard Swedberg. Cambridge: Cambridge University Press: 45-73.

Evans-Pritchard, Edward. 1976 [1937]. *Witchcraft, Oracles, and Magic Among the Azande*. Oxford: Clarendon Press.

Eyre, Dana P., and Mark C. Suchman. 1996. "Status, Norms, and the Proliferation of Conventional Weapons: An Institutional Theory Approach." In *The Culture of National*

Security: Norms and Identity in World Politics, ed. Peter J. Katzenstein. New York: Columbia University Press: 79-114.

Fair, David. 1979. "Causation and the Flow of Energy." *Erkenntnis* 14: 219-50.

Farber, H., and J. Gowa. 1997. "Common Interests or Common Polities?" *Journal of Politics* 57: 393-417.

Fay, Brian. 1994 [1983]. "General Laws and Explaining Human Behavior," In *Readings in the Philosophy of Social Science*, eds. Michael Martin and Lee C. McIntyre. Cambridge, MA: The MIT Press: 91-110.

Fetzer, James H., ed. 1988. *Probability and Causality: Essays in Honor of Wesley C. Salmon*. Boston, Dordrecht.

Fierke, K.M. 2002. "Links Across the Abyss: Language and Logic in International Relations." *International Studies Quarterly* 46: 331-54.

Filson, Darren, and Suzanne Warner. 2004. "Bargaining and Fighting: The Impact of Regime Type on War Onset, Duration, and Outcomes." *American Journal of Political Science* 48: 296-313.

Finnemore, Martha, and Kathryn Sikkink. 2001. "Taking Stock: The Constructivist Research Program in International Relations and Comparative Politics." *Annual Review of Political Science* 4: 391-416.

Forbes, H. Donald. 2004. "Positive Political Theory." In *Handbook of Political Theory*. Eds. Gerald F. Gaus and Chandran Kukathas. London: Sage Publications: 57-72.

Freedman, Lawrence. 1997. "How Did the Democratic Process Affect Britain's Decision to Reoccupy the Falkland Islands?" In *Paths to Peace: Is Democracy the Answer?*, ed. Miriam Fendius Elman. Cambridge, MA: The MIT Press: 235-266.

Frey, R.G. 1976. "Judgments of Causal Importance in the Social Sciences." *Philosophy of Social Science* 6: 245-48.

Friedman, Gil and Harvey Starr. 1997. *Agency, Structure, and International Politics: From Ontology to Empirical Inquiry*. London: Routledge.

Friedman, Jeffrey. 1996. "Economic Approaches to Politics." In *The Rational Choice Controversy*. Ed. Jeffrey Friedman. New Haven, CT: Yale University Press: 1-24.

Frochlic, N. and J. Oppenheimer. 1984. "Beyond Economic Man: Altruism, Egalitarianism, and Difference Maximizing." *Journal of Conflict Resolution* 28:3-24.

Galavotti, Maria Carla and Allesandro Pagnini, eds. 1999. *Experience, Reality, and Scientific Explanation: Essays in Honor of Merrilee and Wesley Salmon*. Boston, MA: Dordrecht.

Garavito, Michael R. 1999. "Working Knowledge: Aspirin." *Scientific American*. 280: 108.

Gartzke, Eric. 1998. "Kant We All Just get Along? Opportunity, Willingness, and the Origins of the Democratic Peace." *American Journal of Political Science* 42: 1-27.

Geertz, Clifford. 1973. *The Interpretation of Cultures*. New York: Basic Books.

Geertz, Clifford. 1980. "Blurred genres: The Refiguration of Social Thought." *American Scholar* 49: 165-79.

Geertz, Clifford. 1994 [1983]. "Thick Description: Toward an Interpretive Theory of Culture", In *Readings in the Philosophy of Social Science*, eds. Michael Martin and Lee C. McIntyre. Cambridge, MA: The MIT Press: 213-31.

George, Alexander and Timothy McKeown. 1985. "Case Studies and Theories of Organizational Decision Making." In *Advances in Information Processing in Organizations, A Research Annual on Public Organizations* Vol. 2, eds Robert F. Coulam and Richard A. Smith. Greenwich, CT: JAI Press: 21-58.

George, Alexander L. and Andrew Bennett. 2005. *Case Studies and Theory Development in the Social Sciences* (Cambridge, MA: MIT Press).

Gerring, John. 2004. "What is a Case Study and What is it Good for?" *American Political Science Review*: 98: 341-54.

Gerring, John. 2005. "Causation: A Unified Framework for the Social Sciences." *Journal of Theoretical Politics* 17: 163-98.

Gleditsch, Nils Petter, and Havard Hegre. 1997. "Peace and Democracy: Three Levels of Analysis." *Journal of Conflict Resolution* 41: 283-310.

Gleditsch, Nils Petter. 2001. "Armed Conflict and the Environment." *Journal of Peace Research* 35: 381-400.

Glennan, Stuart S. 1996. "Mechanisms and the Nature of Causation." *Erkenntnis* 44: 49-71.

Glymour, Clark, R. Scheinners, P. Spirtes, and K. Kelly. 1982. *Discovering Causal Structure: Artificial Intelligence, Philosophy of Science, and Statistical Modeling*. San Diego: Academic Press.

Gowa, Joanne. 1999. *Ballots and Bullets: The Elusive Democratic Peace*. Princeton, NJ: Princeton University Press.

Grady, Denise. 1996. "So, Smoking Causes Cancer: This is News?" *New York Times* October 27, Section 4: 3.

Green, Daniel M. 2002. "Constructivist Comparative Politics: Foundations and Framework," In *Constructivism and Comparative Politics*, ed. Daniel M. Green. London: M.E. Sharpe.

Green, Donald, and Ian Shapiro. 1994. *Pathologies of Rational Choice Theory: A Critique of Applications in Political Science*. New Haven, CT: Yale University Press.

Greenwood, John D. 1994. *Realism, Identity and Emotion: Reclaiming Social Psychology*. London, Sage.

Groopman, Jerome. 2001. "The Thirty Year's War: Have We Been Fighting Cancer the Wrong Way?" *The New Yorker* June 4: 52-63.

Hammond, Michael. 1977. "Weighting Causes in Historical Explanation." *Theoria* 43: 103-28.

Hardin, Garrett. 1968. "The Tragedy of the Commons." *Science* 162:1243-48.

Harré, Rom. 1985. *The Philosophies of Science*. Oxford: Oxford University Press.

Harré, Rom. and E.H. Madden. 1975. *Causal Powers*. Totowa, NJ: Rowman and Littlefield.

Hempel, Carl G. 1959. "The Empiricist Criteria of Meaning" In *Logical Positivism*, ed. A.J. Ayer. Glencoe, IL: The Free Press.

Hempel, Carl G. 1962. "Explanation in Science and in History." In *Frontiers of Science and Philosophy*, ed. R.G. Colodny. Pittsburgh: University of Pittsburgh Press: 9-33.

Hempel, Carl G. 1965. *Aspects of Scientific Explanation*. New York: The Free Press, 1965.

Hoelt, William. 1993. *Explaining Interdemocratic Peace: The Norm of Cooperatively Biased Reciprocity*. Ph.D. dissertation. Georgetown University.

Homer-Dixon, Thomas F. 2000. *The Ingenuity Gap: How Can We Solve the Problems of the Future?* Toronto: Knopf Canada.

Homer-Dixon, Thomas F. 1999. *Environmental Scarcity and Violent Conflict*. Princeton: Princeton University Press.

Honderich, Ted, ed. 1995. *The Oxford Companion to Philosophy*. Oxford: Oxford University Press.

Hopf, Ted. 1998. "The Promise of Constructivism in International Relations Theory." *International Security* 23: 171-200.

Hull, Clark L. 1966 [1943]. *Principles of Behavior: An Introduction to Behavior Theory*. New York, Appleton-Century-Crofts.

Hull, Clark L. 1952. *A Behavior System: An Introduction to Behavior Theory Concerning the Individual Organism*. New Haven: Yale University Press.

Hume, David. 1975 [1748]. *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*. Oxford: Clarendon Press.

Hume, David. 1978 [1739]. *A Treatise of Human Nature*. Oxford: Clarendon Press.

Huth, Paul K. and Todd L. Allee. 2002. *The Democratic Peace and Territorial Conflict in the Twentieth Century*. Ann Arbor: Cambridge University Press.

Jones, Todd. 1998. "Interpretive Social Science and the 'Native's Point of View': A Closer Look." *Philosophy of the Social Sciences* 28: 32-68.

Jungblut, Bernadett M.E., Richard J. Stoll. 2002. "The Liberal Peace and Conflictive Interactions: The Onset of Militarized Interstate Disputes, 1950-78." *Journal of Peace Research* 39: 527-46.

Kahl, Colin H. 1999. "Constructing a Separate Peace: Constructivism, Collective Liberal Identity, and Democratic Peace." *Security Studies* 8: 94-144.

Kasza, Gregory J. 2001. "Perestroika: For an Ecumenical Science of Politics." *PS: Political Science & Politics* 34: 597-99.

Katzenstein, Peter J., ed. 1996a. *The Culture of National Security: Norms and Identity in World Politics*. New York: Columbia University Press.

Katzenstein Peter J., 1996b. "Introduction: Alternative Perspectives on National Security." In *The Culture of National Security: Norms and Identity in World Politics*, ed. Peter J. Katzenstein. New York: Columbia University Press: 1-32.

Keat, R., and J. Urry. 1975. *Social Theory as Science*. London: Routledge & Paul.

Kenny, David A. 1979. *Correlation and Causality*. New York: John Wiley & Sons.

Keohane, Robert. 1999. "Dinosaurs, Detectives and Causal Mechanisms: Coping with Uniqueness in Social Science Research." Unpublished manuscript presented at the American Political Science Association Annual Meeting, September 4, 1999, Atlanta, Georgia.

Kincaid, Harold. 1994 [1990]. "Defending Laws in the Social Sciences." In *Readings in the Philosophy of Social Science*, eds. Michael Martin and Lee C. McIntyre. Cambridge, MA: The MIT Press: 111-31.

King, Gary, Robert Keohane, and Sydney Verba. 1994. *Designing Social Inquiry*. Princeton: Princeton University Press.

Kivimäki, Timo. 2001. "The Long Peace of ASEAN." *Journal of Peace Research* 38: 5-25.

Koslowski, Barbara. 1996. *Theory and Evidence: The Development of Scientific Reasoning*. Cambridge, MA: The MIT Press.

Lalman, David, Joe Oppenheimer, and Piotr Swistak, 1993. "Formal Rational Choice Theory: A Cumulative Science of Politics," In *Political Science: The State of the Discipline II*, ed. Ada W. Finifter. Washington D.C.: The American Political Science Association: 77-104.

Lawson, Tony. 1997. *Economics and Reality*. London: Routledge.

Layne, Christopher. 1994. "Kant or Cant: The Myth of the Democratic Peace."
International Security 19: 5-49.

Leamer, Edward E. 1983. "Let's Take the Con out of Econometrics."
American Economic Review 73: 31-43.

Levy, Jack. 1988. "Domestic Politics and War." *Journal of Interdisciplinary History* 18:
653-73.

Lewontin, Richard C. 1976. "The Analysis of Variance and the Analysis of Causes." In
The IQ Controversy: Critical Readings, eds. N.J. Block and Gerald Dworkin. New York,
NY: Pantheon.

Lijphart, Arend. 1971. "Comparative Politics and the Comparative Method." *American
Political Science Review* 65: 682-93.

Lipset, Seymour. 1960. *Political Man*. Garden City, NY: Doubleday & Company, Inc.

Lipson, Charles. 2003. *Reliable Partners: How Democracies Have Made A Separate
Peace*. Princeton, NJ: Princeton University Press.

- Lipton, Michael. 1992. "Economics and Anthropology: Grounding Models in Relationships." *World Development* 20: 1541-46.
- Little, Daniel. 1998. *Microfoundations, Method, and Causation: On the Philosophy of Social Sciences*. New Brunswick, NJ : Transaction Publishers.
- Little, Daniel. 1991. *Varieties of Social Explanation*. Boulder: Westview Press.
- Louch, Alfred. 1966. *Explanation and Human Action*. Oxford: Basil Blackwell.
- MacDonald, Paul K. 2003. "Useful Fiction or Miracle Maker: The Competing Epistemological Foundations of Rational Choice Theory." *American Political Science Review* 97: 551-65.
- Mackenzie, Brian D. 1977. *Behaviorism and the Limits of the Scientific Method*. London: Routledge.
- MacIntyre, Alasdair. 1962. "A Mistake About Causality in Social Science." In *Philosophy, Politics, and Society (Second Series)*, eds. P. Laslett and W.G. Runciman. Oxford: Basil Blackwell: 48-70.

Mahoney, James. 2001. "Beyond Correlational Analysis: Recent Innovations in Theory and Method." *Sociological Forum* 16: 575-93.

Maoz, Zeev, and Bruce Russett. 1993. "Normative and Structural Causes of Democratic Peace, 1946-1986." *American Political Science Review* 87: 624-38.

Martin, Michael. 1994. "Taylor on Interpretation and the Sciences of Man." In *Readings in the Philosophy of Social Science*, eds. Michael Martin and Lee C. McIntyre. Cambridge, MA: The MIT Press: 264-265.

Martin, Michael, and Lee C. McIntyre. 1994. "Introduction to Part V." In *Readings in the Philosophy of Social Science*, eds. Michael Martin and Lee C. McIntyre. Cambridge, MA: The MIT Press:345-7.

Martin, Steve. 1998. "Studies in the new Causality." *The New Yorker* 74: 108.

McClelland, Peter. 1975. *Explanation and Model Building in History, Economics, and the New Economic History*. Ithica, N.Y., Cornell University Press.

Martin, R. 1972. "On weighting causes." *American Philosophical Quarterly* 9: 291-9.

McKim, Vaughn R. 1997. "Introduction." In *Causality in Crisis? Statistical Methods and the Search for Causal Knowledge in the Social Sciences*, eds. Vaughn R. McKim and Stephen P. Turner. Notre Dame: University of Notre Dame Press: 1-19.

McKim, Vaughn R. and Stephen P. Turner, eds. 1997. *Causality in Crisis? Statistical Methods and the Search for Causal Knowledge in the Social Sciences*. Notre Dame: University of Notre Dame Press.

McLaughlin, Robert, ed. 1982. *What? Where? When? Why? Essays on Induction, Space, Time, and Explanation: Inspired by the Work of Wesley C. Salmon and Celebrating his First Visit to Australia, September-December 1978*. Boston, Dordrecht.

Mellor, D.H. 1974. "In Defense of Dispositions." *Philosophical Review* 83: 157-81.

Mesquida, Christian G. and Neil I. Wiener. 1996. "Human Collective Aggression: A Behavioral Ecology Perspective." *Ethology and Sociobiology* 17: 247-62.

Miller, Richard W. 1987. *Fact and Method: Explanation, Confirmation and Reality in the Natural and Social Sciences*. Princeton, NJ: Princeton University Press.

Mills, John A. 1998. *Control: A History of Behavioral Psychology*. New York: New York University Press.

Mills, John Stuart. 1843. *A System of Logic*. London, Longmans.

Mintz, Alex, and Nehemia Geva. 1993. "Why Don't Democracies Fight Each Other? An Experimental Study." *The Journal of Conflict Resolution* 37: 484-503.

Monroe, Kristen Renwick. 2004. "The Chicago School: Forgotten but Not Gone." *Perspectives on Politics* 2: 94-8.

Moore, John Norton. 2004. *Solving the War Puzzle: Beyond the Democratic Peace*. Durham, NC: Carolina Academic Press.

Nagel, Ernest. 1961. *The Structure of Science: Problems in the Logic of Scientific Explanation*. New York. Harcourt, Brace, and World.

Oneal, John R., Frances H. Oneal, Zeev Maoz, and Bruce Russett. 1996. "The Liberal Peace: Interdependence, Democracy, and International Conflict, 1950-85." *Journal of Peace Research* 33: 11-28.

Oneal, John R., and James Lee Ray. 1997. "New Tests of the Democratic Peace: Controlling for Economic Interdependence, 1950-85." *Political Research Quarterly* 50: 751-75.

Outhwaite, W. 1975. *Understanding Social Life. The Method Called Verstehen*. London: George Allen & Unwin Ltd.

Owen, John M. 1994. "How Liberalism Produces Democratic Peace." *International Security* 19:87-125.

Paolo, Vineis and Miquel Porta. 1996. "Causal Thinking, Biomarkers, and Mechanisms of Carcinogenesis." *Journal of Clinical Epidemiology* 49: 951-6.

Parker, John, Leonard Mars, Paul Ransome, and Hilary Stanworth. 2003. *Social Theory: A Basic Tool Kit*. Houndmills: Palgrave MacMillan.

Pawson, Ray and Nick Tilley. 1997. *Realistic Evaluation*. London: Sage Publications.

Peattie, Lisa. 1995. "An Approach to Urban Research in the 1990s." In *Perspectives on the City Vol. 4*, eds. Richard Stren and Judith Kjellberg Bell. Toronto: University of Toronto Press.

Poling, A., L.L. Methot, and M.G. LeSage. 1995. *Fundamentals of Behavior Analytic Research*. New York: Plenum.

Popper, Karl. 1965. *Conjectures and Refutations: The Growth of Scientific Knowledge*. New York. Basic Books Inc., Publishers.

Popper, Karl. 1979. "Three Worlds." *The Michigan Quarterly Review* 18: 1-23.

Pork, Anrus. 1985. "Assessing Relative Causal Importance in History." *History and Theory* 24: 62-9.

Przeworski, Adam and Henry Teune. 1982. *The Logic of Comparative Social Inquiry*.
Malabar, FL: Robert E. Krieger Publishing Company, Inc.

Price, Richard, and Nina Tannenwald. 1996. "Norms and Deterrence: The Nuclear and Chemical Weapons Taboos." In *The Culture of National Security: Norms and Identity in World Politics*, ed. Peter J. Katzenstein. New York: Columbia University Press: 114-52..

Psillios, Stathis. 1999. *How Science Tracks Truth*. London: Routledge.

Putnam, Robert D. 1993. *Making Democracy Work: Civic Traditions in Modern Italy*.
Princeton, NJ: Princeton University Press.

Quine, Willard Van Orman. 1961. *From a Logical Point of View*. Cambridge, MA:
Harvard University Press.

Ragin, Charles. 1987. *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies*. Berkeley, CA: University of California Press.

- Raknerud, Arvid, and Havard Hegre. 1997. "The Hazard of War: Reassessing the Evidence for the Democratic Peace." *Journal of Peace Research* 34: 385-404.
- Ray, James Lee. 1995. *Democracy and International Conflict: An Evaluation of the Democratic Peace Proposition*. Columbia, SC: University of South Carolina Press.
- Reed, William. 2003. "Information, Power, and War." *American Political Science Review* 97: 633-41.
- Remmer, Karen L. 1998. "Does Democracy Promote Interstate Cooperation? Lessons from the Mercorsur Region." *International Studies Quarterly* 42: 25-52.
- Risse-Kappen, Thomas. 1995. "Democratic Peace – Warlike Democracies? A Social Constructivist Interpretation of the Liberal Argument." *European Journal of International Relations* 1: 491-517.
- Roberts, Clayton. 1996. *The Logic of Historical Explanation*. University Park, PA: The Pennsylvania State University Press.
- Rogowski, Ronald. 1995. "The Role of Theory and Anomaly in Social-Scientific Inference." *American Political Science Review* 89: 467-70.

- Ron, Amit. 2002. "Regression Analysis and the Philosophy of Social Sciences – a Critical Realist View." *Journal of Critical Realism* 1: 115-36.
- Rosato, Sebastian. 2003. "The Flawed Logic of Democratic Peace Theory." *American Political Science Review* 97: 585-602.
- Rousseau, David L., Christopher Gedi, Dan Reiter, and Paul K. Huth. 1996. "Assessing the Dyadic Nature of the Democratic Peace, 1918-88." *American Political Science Review* 90: 512-33.
- Ruggie, John Gerard. 1998. "What Makes the World Hang Together? Neo-utilitarianism and the Social Constructivist Challenge." *International Organization* 52: 855-85.
- Rummel, R.J. 1995. "Democracies ARE Less Warlike Than Other Regimes." *European Journal of International Relations* 1: 457-79.
- Runkle, Benjamin Gordon. 2003. *Symptoms or Disease? Arms Races and the Causes of War*. Ph.D. Thesis. Harvard University.
- Russett, Bruce and John Oneal. 2001. *Triangulating Peace: Democracy, Interdependence, and International Organizations*. New York: W.W. Norton & Company.

Salmon, Wesley. 1977. "An At-At theory of Causal Influence." *Philosophy of Science* 44: 215-24.

Salmon, Wesley. 1984. *Scientific Explanation and the Causal Structure of the World*. Princeton, NJ: Princeton University Press.

Salmon, Wesley. 1998. *Causality and Explanation*. New York, NY: Oxford University Press.

Sambanis, Nicholas. 2004. "Using Case Studies to Expand Economic Models of Civil War." *Perspectives on Politics* 2: 259-79.

Sartori, Giovanni, ed. 1984. *Social Science Concepts: A Systematic Analysis*. Beverly Hill: Sage Publications.

Sartori, Giovanni. 1970. "Concept Misinformation in Comparative Politics." *The American Political Science Review* 64: 1033-53.

Sayer, Andrew. 2000. *Realism and Social Science*. London: Sage.

Schultz, Kenneth A. 2001. *Democracy and Coercive Diplomacy*. Cambridge: Cambridge University Press.

Schwartz, Daniel M., Tom Deligiannis, and Thomas Homer-Dixon. 2001. "The Environment and Violent Conflict." In *Environmental Conflict*, eds. Paul F. Diehl and Nils Petter Gleditsch. Boulder, CO: Westview Press: 273-94.

Schwartz, Thomas and Kiron K. Skinner. 2002. "The Myth of the Democratic Peace." *Orbis* 46: 159-72.

Scriven, Michael. 1976. "Maximizing the Power of Causal Investigations: The Modus Operandi Method." In *Evaluation Studies 1*, ed. Gene V. Glass. London: Sage Publications: 101-18.

Scriven, Michael. 1962. "Explanations, Predictions, and Laws." In *Scientific Explanation, Space & Time*, eds. Herbert Feigl and Gover Maxwell. Minnesota Studies in the Philosophy of Science 3: 170-230. Minneapolis: University of Minnesota Press.

Searle, John. 1995. *The Construction of Social Reality*. New York: The Free Press.

Shaffer, Paul. 2002. "Participatory Analyses of Poverty Dynamics: Reflections on the Myanmar PPA." In *Knowing Poverty: Critical Reflections on Participatory Research and Policy*, eds. K Brock and R McGee. London: EarthEcan Publications Ltd.

Skinner, B.F. 1974. *About Behaviorism*. New York, NY: Alfred A. Knopf.

Skocpol, Theda. 1994. *Social Revolutions in the Modern World*. Cambridge: Cambridge University Press.

Small, Melvin, and J. David Singer. 1976. "The War Proneness of Democratic Regimes." *Jerusalem Journal of International Relations* 1: 50-69.

Snyder, J. 2000. *From Voting to Violence*. New York, NY: W.W. Norton.

Sober, Elliott. 1988. "Apportioning Causal Responsibility." *The Journal of Philosophy* 85: 303-18.

Specter, Michael. 2001. "Rethinking the Brain: How the Songs of Canaries Upset a Fundamental Principle of Science." *The New Yorker* July 23: 42-53.

Spiro, David E. 1994. "The Insignificance of the Liberal Peace." *International Security* 19: 50-86.

Starr, Harvey. 1997. "Democracy and Integration: Why Democracies Don't Fight Each Other." *Journal of Peace Research* 34: 153-62.

Strawson, P. F. 1985. "Causation and Explanation," In *Essays on Davidson: Actions and Events*, eds. B. Vermazen and M. Hintikka. Oxford: Clarendon Press.

Suppes, Patrick. 1970. *A Probabilistic Theory of Causality*. Amsterdam: North Holland Publishing Company.

Tarrow, Sydney. 1995. "Bridging the Quantitative-Qualitative Divide in Political Science." *American Political Science Review* 89: 471-4.

Taylor, Charles. 1985. *Philosophy and the Human Sciences (Philosophical Thesiss, Vol. 2)*. Cambridge: Cambridge University Press.

Taylor, Charles. 1994 [1971]. "Interpretation and the Sciences of Man." In *Readings in the Philosophy of Social Science*, eds. Michael Martin and Lee C. McIntyre. Cambridge, MA: The MIT Press: 181-212.

Tetlock, Philip E. and Aaron Belkin. 1996. *Counterfactual Thought Experiments in World Politics: Logical, Methodological, and Psychological Perspectives*. Princeton, NJ: Princeton University Press.

Thagard, Paul. 1999. *How Scientists Explain Disease*. Princeton: Princeton University Press.

Thompson, W.R., and R. Tucker. 1997. "A Tale of Two Democratic Peace Critiques." *Journal of Conflict Resolution* 41: 428-54.

Thyer, Bruce A. 1999. *The Philosophical Legacy of Behaviorism*. Dordrecht: Kluwer.

Tilly, Charles. 1997. "Means and Ends of Comparison in Microsociology." *Comparative Social Research* 16: 43-53.

Tolman, E. C. 1932. *Purposive Behavior in Animals and Men*. New York: Century.

Uttal, William R. 2000. *The War Between Mentalism and Behaviorism*. Mahwah, NJ: Lawrence Erlbaum Associates.

Van Fraassen, Bas C. 1980. *The Scientific Image*. Oxford: Clarendon Press.

Von Wright, Georg Henrik. 1971. *Explanation and Understanding*. Ithica, NY: Cornell University Press.

Walt, Stephen M. 1999. "Rigor or Rigor Mortis? Rational Choice and Security Studies." *International Security* 23: 5-48.

Waltz, Kenneth. 1979. *Theory of International Politics*. New York, NY: McGraw-Hill.

Waltz, Kenneth and Scott Sagan. 2002. *The Spread of Nuclear Weapons: A Debate Renewed*. Stanford: Stanford University Press.

Wantchekon, Leonard. 2004. "The Paradox of 'Warlord' Democracy: A Theoretical Investigation." *American Political Science Review* 98: 17-33.

Watson, John B. 1966 [1924]. *Behaviorism*. Chicago. Chicago University Press.

Weber, Max. 1968. *Economy and Society: An Outline of Interpretive Sociology*. New York: Bedminster Press.

Weber, Max. 1949. *The Methodology of the Social Sciences*. New York: The Free Press.

Wendt, Alexander. 1999. *Social Theory of International Politics*. Cambridge: Cambridge University Press.

Wilber, Charles K. and Robert S. Harrison. 1978. "The Methodological Basis of Institutional Economics: Pattern Model, Storytelling, and Holism." *Journal of Economic Issues* 12: 61-89.

Williams, Michael C. 2001. "The Discipline of the Democratic Peace: Kant, Liberalism and the Social Construction of Security Communities." *European Journal of International Relations* 7: 525-33.

Winch, Peter. 1958. *The Idea of a Social Science*. London: Routledge and Kegan Paul.

Windmaier, Wesley W. 2004. "The Social Construction of the 'Impossible Trinity': The Intersubjective Bases of Monetary Cooperation." *International Studies Quarterly* 433-53.

Yee, Albert S. 1996. "The Causal Effects of Ideas on Policies." *International Organization* 50: 69-108.

Figure 1: Cartwright's Causal Arrangement

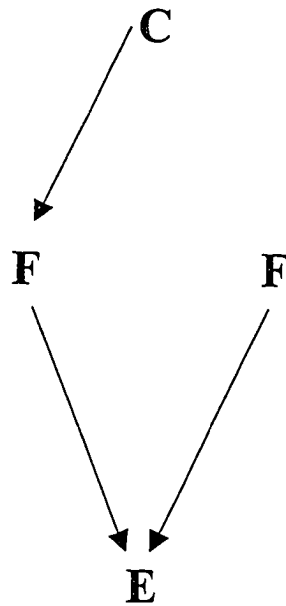


Table 1: Ontological and Epistemological Breakdown of Mechanisms

Ontological Property	Method to Identify Ontological Property	Manner by which Method Generates Findings to Explain Regularities
Physicalness	Process-Tracing	As a chain of events/phenomena
Agency	Rational Choice Modeling	As the aggregation or strategic interaction of individual choices
Intentionality	Interpretation	As “enabling” intersubjective meanings

Table 2: Definitions of Mechanism From Mahoney (2001, 579-80)

Bennett and George: “[T]he processes and intervening variables through which causal or explanatory variables produce causal effects.”

Bhaskar: “[A] model...which *if* it were to exist and act in the postulated way would account for the phenomenon in question.”

Boudon: “[T]he well-articulated set of causes responsible for a given social phenomenon.”

Cowen: “[R]ational-choice accounts of how a specified combination of preferences and constraints can give rise to more complex social outcomes.”

Elster: “[N]uts and bolts, cogs and wheels – that can be used to explain quite complex social phenomena.”

Gambetta: “[H]ypothetical causal models that make sense of *individual behavior*. ”

Goldthorpe: “[S]ome process existing in time and space, even if not perhaps directly observable, that actually generates the causal effect of X and Y and, in doing so, produces the statistical relationship that is empirically in evidence.”

Harré: “[T]he structures, states, and inner constitutions from which the phenomena of nature flow.”

Hedström and Swedberg: “[S]ome form of ‘causal agent’ that is assumed to have generated the relationship between the entities observed.”

Hernes: “[A] set of interacting parts...an abstract, dynamic logic by which social scientists render understandable the reality they depict.”

Keat and Urry: “[T]he ‘nature,’ ‘essence,’ or ‘inner constitution’ of various types of entity.”

Kiser and Hechter: “[T]he process by which one variable influences the other, in other words, how it is that X produces Y.

Koslowski: “[A] theory or an explanation, and what it explains is how one event causes another.”

Little: “[A] series of events governed by lawlike regularities that lead from the explanans to the explanandum.”

Schelling: “[A] plausible hypothesis, or set of plausible hypotheses, that could be the explanation of some social phenomenon, the explanation being in terms of interactions between individuals and other individuals, or between individuals and some social aggregate.”

Sørensen: “[A]n account of how change in some variable is brought about – a conceptualization of what ‘goes into’ a process.”

Steinmetz: “[T]endencies’ rather than ‘powers’ because they are not just potentialities but potentialities that may be exercised with being manifested.”

Stinchcombe: “[B]its of ‘sometimes true theory’ or ‘model’ that represent a causal process, that have some actual or possible empirical support separate from the larger theory in which it is a mechanism, and generate increased precision, power, or elegance in the large-scale theories.”

Tilly: “...selective explanation of salient features by means of partial analogy....Mechanisms are events that alter relations among some specified set of elements.”

Table 3: Four Philosophical Positions on Causation

Epistemological Status of Theoretical Mechanisms	Ontological Status of Unobservables		
		Fictional	Literally True or False
	Irrelevant to Scientific Inquiry	<p>1</p> <p>Humean Empiricism (HE)</p> <p>Causation is <i>not</i> genuinely meaningful: unobservables are <i>useless</i> fictions; mechanisms are irrelevant because empirical regularities cannot be explained</p>	<p>3</p> <p>Humean Realism (HR)</p> <p>Causation is <i>not</i> genuinely meaningful: unobservables have a truth value; but mechanisms are irrelevant because even unobservable process are based on regularities</p>
Relevant to Scientific Inquiry	<p>2</p> <p>Instrumentalism Empiricism (IE)</p> <p>Causation is <i>not</i> genuinely meaningful: unobservables are <i>useful</i> fictions; mechanisms are relevant because they can help predict but have no genuine explanatory power</p>	<p>4</p> <p>Scientific Realism (SR)</p> <p>Causation <i>is</i> genuinely meaningful: unobservables have a truth value; mechanisms are relevant because they help predict and retrodict and have genuine explanatory power</p>	