Uncertainties in Models and Empirics

Desiree A. Desierto*

April 14, 2021

Abstract

How much knowledge can social science research generate? I formulate lower bounds of uncertainty in models and empirical tests of the causal mechanism behind a social phenomenon. I apply the framework to the study of status inequality.

1 Introduction

How much do we really learn from social science research?

Consider the standard approach to the causal analysis of a social phenomenon. One formulates, implicitly or explicitly, a verbal or formal model of the phenomenon, from which propositions or hypotheses can be logically derived. These hypotheses are then taken to the data to be falsified. What, then, can be inferred? When falsification fails, the only thing that can be concluded is that the underlying *model* is plausible. Yet the model is not necessarily the true causal mechanism that gives rise to the phenomenon. It is simply one of many – possibly infinite (for how do we know the number?), mechanisms that can generate the phenomenon.[1]

But, with sufficiently many occurrences of the phenomenon, for which the same empirical test fails to falsify the same model — does this not improve our confidence that the model closely approximates the true causal mechanism?

*George Mason University, Economics Department, and University of Maryland, School of Public Policy; ddesiert@gmu.edu

[1]Skarbek (2020) notes the 'equifinality' problem in social science methodology in which "there might be multiple possible causal pathways that connect x to y". See also Gerring (2012).

Herein lies the problem, for such multiple occurrences may not exist. This is because a social phenomenon is a construct with no fixed definition, and therefore cannot be replicated or replicable with certainty. Unlike in the physical sciences, social phenomena cannot be treated as physical matter whose definition is exact and does not depend on individuals' conceptual notions. The terms 'state' or 'autocrat' or 'social inequality' can mean different things to different people studying them from different points in time, whereas light is, everywhere and always, electromagnetic radiation that can be seen by the human eye.

This subjectivity of social concepts, and the difficulty of operationalizing such concepts to render it suitable for scientific analysis, has long been pointed out by social scientists. Hayek (1952) was probably the first to articulate the problem when he noted the problems that arise when social scientists adopted the methods of the physical science.[2] The problem of concept operationalization certainly has not gone away — modern social science methodology textbooks actually devote chapters to the issue (e.g. King, Keohane and Verba (1994) and Gerring (2012)). Coppedge (1999) argues that "thick concepts and theories are unwieldy in generalizing or rigorously testing complex hypotheses".

What, then, do social phenomena have in common with physical phenomena, which renders the former suitable for scientific study? The fundamental similarity, in my view, is that both types of phenomena occur in time. Every antecedent is followed by a consequence which is also an antecedent for subsequent events.[3] To uncover the causal mechanism that produces the phenomenon is thus to correctly identify the chain of events that lead to the phenomenon, and the manner by which they do so. This

[2] Following Hayek, economists in the modern Austrian tradition have argued for the use of rich descriptive, anthropological, and historical analysis in conjunction with economic theory (see Boettke, 1997, 2000; Coyne and Boettke, 2015; Martin, 2015).

[3] The events that are relevant to social phenomena are those involving human action (von Mises, 1949, 1996), whereas physical phenomena may occur independently of humans.

is what makes research in the physical sciences useful – the fact that time follows a linear path, which allows a cause-and-effect analysis, and this similarly enables causal analysis in the social sciences.[4]

The difference, however, is that uncovering the causal chain underlying a social phenomenon is wrought with far more uncertainties, beginning with the definition and operationalization of the phenomenon. There is also the uncertainty from the construction of a model that can explain the emergence of the phenomenon, and the uncertainty as to whether it approximates the true causal mechanism. There is the uncertainty in the design of empirical tests, and the uncertainty as to the extent that an empirical result validates the model.

Other papers have indeed pointed out such uncertainties. Duhem (1906) and Quine (1951) were the first to elucidate indeterminacies in the link between theory and data. This is encapsulated in their Duhem-Quine thesis by which "one can never be sure that it is a given theory rather than auxiliary or background hypotheses which experiment has falsified" (Harding, 1976). Thus, "theories can be submitted to test only in conjunction with a set of assumptions and rules of inference. These constitute the auxiliary hypotheses and include the simplifying assumptions of the theory, the calibration of the experiment and the axioms of statistical inference used" (Sawyer et al., 1997). More recently, social scientists have either focused on uncertainties underlying (empirical) causal inference (e.g. Ruhm (2019), Deaton and Cartwright (2018)), or on what we can actually learn from models (e.g. Clarke and Primo (2011)).

To the best of my knowledge, this paper is the first to combine all possible sources of uncertainty – from concept operationalization to model construction and empirical

[4] Jaki demonstrates that in the history of science, wherever and whenever a circular view of time is adopted – the so-called "eternal returns" view of the universe, research in the physical sciences experience a 'still birth' ' (Jaki, 1974, 1978, 2002).

testing, in order to provide a simple formulation of the lower bounds of uncertainty that can be applied to causal analysis of social phenomena. Section 2 presents the framework, and section 3 applies it to a study of social inequality. Section 4 concludes with some reflections on the value of generalizability of findings in social science research.

2 An Uncertainty Principle

Suppose a researcher is interested in finding out the cause of some social phenomenon. After some preliminary analysis of different kinds and sources of information, she may then propose a verbal or formal model of the causal mechanism that gives rise to the phenomenon, or provide some empirical evidence of the mechanism, or both. How much knowledge can be gained from such endeavor? Very little, as it turns out.

This is not a blind assertion. In fact, to be precise, I formulate the lower bounds of uncertainty that not even the best causal analysis of a social phenomenon can surmount. The framework is useful in comparing the extent of knowledge one can gain from different studies of the same phenomenon, as well as comparing across different approaches – theoretical, empirical, or a combination thereof.

To proceed, suppose one wanted to explain how a particular social phenomenon occurred at time T. Since social phenomena are a result of combinations of human action occurring within a possibly evolving environment, and all such actions occur in time, then one can denote an event E_t as capturing a particular 'step' or set of actions at time t prior to T that contribute to the occurrence of the social phenomenon. Since there are many concurrent events for each t, and possibly many other social phenomena simultaneously occurring at T, the task of correctly identifying each particular event prior to T that contributes to the occurrence of the social phenomenon and the mechanism by which such events bring it about is a non-trivial endeavor, which in fact increases in complexity the greater the number of periods prior to T and the greater the number of concurrent events at each of these prior periods.

Even assuming that the number of relevant prior periods and that the number of events per period are known, uncovering the causal chain from prior events to the social phenomenon of interest is inherently wrought with much uncertainty.

To formally demonstrate, denote such causal chain as C_{ST}. Then C_{ST} is a mapping from the entire space of events that have occurred since d periods prior to T to the set of all social phenomena at T. That is,

$$C_{ST} : \{E_{T-d}\} \times \{E_{T-d+1}\} \times \{E_{T-d+2}\} \times ... \times \{E_{T-d+d}\} \equiv \prod_i \{E_{T-d+i}\} \to \{S_T\},$$

with $i = 0, 1, 2, ..., d$, and where $\{E_{T-d+i}\}$ is the set of all events that occurred in the $|-d+i|$th period prior to T, and $\{S_T\}$ the set of all social phenomena 'created' at T.[5]

This causal chain C_{ST} is unknowable. That is, one cannot observe nor infer the exact mapping from the space of prior events to the particular social phenomenon of interest. Otherwise, if one had this function, then one could just deduce the particular events that caused the particular social phenomenon. Because the mapping C_{ST} is unknowable, any logic that attempts to link the space of events to the phenomenon is necessarily inductive. That is, from the space of events, one can form an idea about C_{ST} that would generate the phenomenon. It is this idea that we call a ' model' or 'theory', which we can denote as m_{ST}. That is, m_{ST} is itself a causal chain which we think can map the space of events to the phenomenon of interest. Given m_{ST}, one can then deduce the particular events that have 'caused' the phenomenon.

Thus, forming a theory or model about how some social phenomenon came about inherently uses inductive logic, and it is only after taking the model as given that one

[5]The set $\{S_T\}$ of social phenomena occurring at T is also a set of antecedents for other events occurring at $T + 1$. Thus, elements in $\{S_T\}$ are also events, but to distinguish them from its antecedents, we let the latter sets be denoted by $\{E_{T-d+i}\}$. This is with some abuse of notation since, when $i = d$, then $\{E_T\}$, which we interpret as the set of direct antecedents to $\{S_T\}$ and, therefore, occurs just before $\{S_T\}$ even though it is subscripted by terminal time T.

can deductively infer the causes of the phenomenon. Note that we have no way of knowing whether m_{ST} is the correct causal chain C_{ST}. In fact, m_{ST} is only one of many functions that can map the space of events into the phenomenon of interest. The number of such mappings, and therefore the uncertainty of whether m_{ST} captures the true causal chain C_{ST} increases with the dimension of the space of events and the number of social phenomena that simultaneously occur at T. More formally:

Proposition 1 *The probability p_m that $m_{ST} \equiv C_{ST}$ decreases with the time d at which the causal chain begins, each dimension $\dim\{E_{T-d+i}\}$, $i = 0, 1, 2, .., d$, and the dimension $\dim\{S_T\}$ of $\{S_T\}$.*

I show the proof here as it is subsequently used to establish the main results, i.e. Propositions 2 and 3. To construct a proof of Proposition 1, start by letting the outcome/phenomenon of interest be $S_{T1} \in \{S_T\}$. This is actually a simplifying assumption — if the phenomenon is a social construct, it can have a subjective definition. Thus, there could be uncertainties as to what S_{T1} really is, or what the elements in $\{S_T\}$ are, if they can even be listed and are therefore knowable. Suppose we abstract from this issue, and take S_{T1} as given. Then a model m_{ST} is only one of many possible mappings from the space $\prod_i \{E_{T-d+i}\}$ into $\{S_T\}$ that produce S_{T1}. How many such mappings are there?

We know S_{T1}, but we do not know the true events that have caused it, nor the mechanism by which each event contributes to its occurrence. However, since events occur in time and, thus, follow a sequence from $T - d$ to T, we know that event/s in, e.g. $\{E_{T-5}\}$ cannot directly affect event/s in $\{E_{T-1}\}$, nor any phenomenon in $\{S_T\}$ for that matter. Instead, the causal chain would have to pick out the event/s in $\{E_{T-5}\}$ that affect event/s in $\{E_{T-4}\}$ which affect event/s in $\{E_{T-3}\}$, and so on, until the direct antecedents $\{E_T\}$ which finally affect element/s in $\{S_T\}$. Thus, one source of uncertainty is that we do not know which event/s from $\{E_{T-d+i}\}$, for each $i = 0, 1, 2, ..., d$, contribute/s to the occurrence of event/s in $\{E_{T-d+i+1}\}$ for $i < d$, nor the event/s in $\{E_T\}$ that contribute/s to the occurrence of $S_{T1} \in \{S_T\}$. Suppose we add another simplifying assumption. That is, let the set of events $\{E_{T-d+i}\}$ be sufficiently restrictive such that only one event from $\{E_{T-d+i}\}$ contributes to one event in $\{E_{T-d+i+1}\}$, with each event equally likely to do so.[6] That is, the causal chain involves an event occurring one at a time until the occurrence of S_{T1}. Then the probability that m_{ST} picks out the true event from $\{E_{T-d+i}\}$ for each i is $\frac{1}{\dim\{E_{T-d+i}\}}$. Thus, the joint probability that m_{ST} picks out all events leading up to S_{T1} that together contribute to the latter's occurrence is the product $\prod_i \frac{1}{\dim\{E_{T-d+i}\}}$.

Next, even if the model m_{ST} correctly picks out all the relevant events leading up to S_{T1}, it might not correctly capture the process by which the events give rise to S_{T1}. The probability that m_{ST} picks up the true and entire mechanism that links each event that contributes to the occurrence of S_{T1} actually depends on two things.

First, there is the possibility that simultaneous phenomena at T, i.e. $S_{Tj} \in \{S_T\}$, $j \neq 1$, themselves map on to the phenomenon of interest S_{T1}. The number of such mappings is a subset of the number of relations among all the elements in $\{S_T\}$ which, in turn, is the sum of all binary relations among them, relations among any three elements, among any four, and so on. Since this is a sum of combinations $C(\dim\{S_T\}, 2), C(\dim\{S_T\}, 3), ..., C(\dim\{S_T\}, \dim\{S_T\})$, then the number of relations among $\{S_T\}$ is a function $f(\dim\{S_T\})$ that is increasing in $\dim\{S_T\}$. Specifically,

$$f(\dim\{S_T\}) = \frac{\dim\{S_T\}}{(\dim\{S_T\}-2)!2!} + \frac{\dim\{S_T\}}{(\dim\{S_T\}-3)!3!} + ... + \frac{\dim\{S_T\}}{(\dim\{S_T\}-\dim\{S_T\})!\dim\{S_T\}!},$$

[6] This is akin to saying that, a priori, we have picked out only the events that are most likely to be relevant to such a level of 'accuracy' that we have no other information as to which of these events are more likely to be relevant than the rest. Thus, our belief is that all events in each of the sets have equal probability of being the true relevant event from the set.

where the last term is equal to one, while each term except the last is increasing in $\dim\{S_T\}$.

Thus, since the number S of mappings from $\{S_{Tj}\}$ to S_{T1} is a subset of the number of relations $f(\dim\{S_T\})$ among elements in $\{S_T\}$, then S is bounded by $f(\dim\{S_T\})$ from above. That is, $S = \mathcal{O}(f(\dim\{S_T\})$ as $\dim\{S_T\} \to \infty$. With each mapping equally likely to be the true one or, equivalently, without prior knowledge that one mapping is more likely than the rest, the probability that the model m_{ST} picks out the correct mapping from S_{Tj} to S_{T1} is $\frac{1}{S}$.

Second, for each $\{E_{T-d+i}\}$, there is the analogous possibility that other events in $\{E_{T-d+i}\}$ map on to the relevant event in $\{E_{T-d+i}\}$ that affects the relevant event in $\{E_{T-d+i+1}\}$. In this case the mechanism that links the relevant event in $\{E_{T-d+i}\}$ to the relevant event in $\{E_{T-d+i+1}\}$ includes the indirect effect of the other elements in $\{E_{T-d+i}\}$ on the relevant event in $\{E_{T-d+i}\}$. The set of all such possible mappings is a subset of all the possible relations among the elements in $\{E_{T-d+i}\}$. The number of such relations is the sum of all binary relations among elements in $\{E_{T-d+i}\}$, relations among any of its three elements, and so on. Thus, it is a function $g(\dim\{E_{T-d+i}\}$ that is increasing in $\dim\{E_{T-d+i}\}$. Specifically,

$$\begin{aligned} g(\dim\{E_{T+d-i}\}) = {} & \frac{\dim\{E_{T+d-i}\}}{(\dim\{E_{T+d-i}\}-2)!2!} + \frac{\dim\{E_{T+d-i}\}}{(\dim\{E_{T+d-i}\}-3)!3!} + \ldots \\ & + \frac{\dim\{E_{T+d-i}\}}{(\dim\{E_{T+d-i}\}-\dim\{E_{T+d-i}\})!\dim\{E_{T+d-i}\}!}, \end{aligned}$$

where the last term is equal to one, while each term except the last is increasing in $\dim\{E_{T+d-i}\}$.

Thus, for each set $\{E_{T+d-i}\}$ of events, the number E_i of mappings from elements in $\{E_{T+d-i}\}$ to the event in $\{E_{T+d-i}\}$ that gives rise to the relevant event in $\{E_{T+d-i+1}\}$ is bounded from above by $g(\dim\{E_{T+d-i}\})$. That is, $E_i = \mathcal{O}(g(\dim\{E_{T+d-i}\})$ as $\dim\{E_{T+d-i}\} \to \infty$. With all such mappings equally likely to be the true mapping for each i, the probability that model m_{ST} captures the true mapping for each i is $\frac{1}{E_i}$.

Putting it all together, the probability p_m that m_{ST} is the true causal chain C_{ST} is the probability that each event, simultaneous or prior to S_{T1}, that contributes to the occurrence of S_{T1}, either directly or indirectly, is included, and that the entire mechanism by which they exert their influence is captured. More specifically:

$$p_m = \Big[\prod_i \Big(\frac{1}{\dim \{E_{T+d-i}\}} \frac{1}{E_i} \Big) \Big] \frac{1}{S},$$

where recall that $\frac{1}{\dim \{E_{T+d-i}\}}$ is the probability that the true event in $\{E_{T+d-i}\}$ is picked up by model m_{ST}; $\frac{1}{E_i}$ is the probability that any effects of the other concurrent events in $\{E_{T+d-i}\}$ on the relevant event in $\{E_{T+d-i}\}$ that contributes to the relevant event in $\{E_{T+d-i+1}\}$ is captured; and $\frac{1}{S}$ the probability that any effects of other events in $\{S_T\}$ on S_{T1} is captured.

Since $\frac{1}{S}$ is decreasing in $\dim \{S_T\}$, $\frac{1}{E_i}$ decreasing in $\dim \{E_{T+d-i}\}$, and $\prod_i(\cdot)$ decreasing in d, this concludes the proof of Proposition 1.

From this, the uncertainty of whether m_{ST} indeed captures the true causal chain C_{ST} is simply $1 - p_m$. Note that this is a lower bound, as it has been constructed with many simplifying assumptions. The uncertainty is typically higher. First of all, the number d of relevant periods prior to the occurrence of the social phenomenon of interest – the beginning of the causal chain, is likely unknown, as are the number of concurrent events in each period $T + d - i$ that are equally likely to have influenced the phenomenon, i.e. the cardinality of $\{E_{T+d-i}\}$. Second, I have assumed no uncertainty in identifying the phenomenon of interest S_{T1} from the set of all simultaneous phenomena $\{S_T\}$ at T. Third, I have also assumed that one and only event per time period is directly involved in the causal chain, whereas there could be multiple such events.

Thus, define lower bound $\underline{m_u} \equiv 1 - p_m = 1 - \Big[\prod_i \Big(\frac{1}{\dim \{E_{T+d-i}\}} \frac{1}{E_i} \Big) \Big] \frac{1}{S}$. The following result is immediate:

Proposition 2 *UNCERTAINTY IN MODELS*

The uncertainty m_u that a model m_{ST} explains social phenomenon of interest S_{T1} is no less than $\underline{m_u}$. That is, $m_u \geq \underline{m_u}$.

What about the uncertainty in empirical exercises that attempt to (statistically) identify the cause of phenomenon S_{T1}? Such exercises test predictions of an underlying model, whether or not the model is explicitly articulated. Suppose the underlying is m_{ST}. Because the predictions can be logically deduced from m_{ST}, there is thus no uncertainty in generating them. However, the operationalization of a prediction into a hypothesis that can be tested using data produces uncertainty — denote this as o_u. In addition, the choice of population from which the statistical sample is drawn produces uncertainty. Whether the chosen population is actually the population that experiences the whole causal chain C_{ST} is unknowable since C_{ST} is unknowable in the first place. Since we have only an idea m_{ST} as to what that causal chain is, the uncertainty in choosing the correct population increases with model uncertainty m_u. Thus, denote such uncertainty as $p_u(m_u)$. Lastly, there is uncertainty e_u from the choice of empirical identification strategy, i.e. from the choice of sample, to statistical methods of estimating parameters.

All these uncertainties cumulate. Even a well-identified empirical analysis which has the lowest possible uncertainty e_u and o_u, denoted as $\underline{e_u}$ and $\underline{o_u}$, respectively, can tell us very little about the entire causal explanation behind S_{T1}, if the uncertainty about the population $p_u(m_u)$ is large. Define the lower bound uncertainty for a well-identified empirical exercise as $\underline{c_u} \equiv \underline{o_u} \cdot p_u(m_u) \cdot \underline{e_u}$. Then:

Proposition 3 *UNCERTAINTY IN EMPIRICAL TESTS*

The uncertainty c_u that the set of results from an empirical test explains the social phenomenon of interest S_{T1} is no less than $\underline{c_u}$. That is, $c_u \geq \underline{c_u}$.

Propositions 2 and 3 imply that, *in general*, an analysis that combines a model m_{ST} and a well-identified empirical test produces less uncertainty about the true casual chain C_{ST} behind phenomenon S_{T1}, than model m_{ST} alone. This is because when e_u and o_u are kept to a minimum, then the total uncertainty is kept at $\underline{c_u} = \underline{o_u} \cdot p_u(m_u) \cdot \underline{e_u}$ which is generally less than m_u. It is possible, however, that a badly designed empirical test adds nothing to our idea about C_{ST}, or can even make the analysis more confusing. If e_u and o_u are large, and $p'(m_u) > 1$, then it is possible that $c_u = o_u \cdot p_u(m_u) \cdot e_u$ is greater than m_u, in which case the model alone is better at explaining S_{T1}.

It is important to note the importance of correctly identifying the relevant population. To keep $p_u(m_u)$ to a minimum, one has to be fairly certain that the population from which the sample is drawn for empirical tests is the population that actually experienced the causal chain C_{ST}. There are two ways by which $p_u(m_u)$ can be otherwise large. First, the uncertainty m_u about whether the model captures the true causal chain can be large. Second, larger m_u may itself generate larger uncertainty as to whether the chosen population is the correct one, i.e. $p'(m_u)$ can be large. These issues may be more relevant when explaining historical social phenomena, as the ability to identify and locate the actual entities that experienced a past phenomenon is itself dependent on the quality of historical accounts and amount of historical detail that are currently available. The use of experimental, rather than observational, data in such cases thus calls attention to the extent to which the true population that experienced the phenomenon is approximated. Do field experiments necessarily produce less *population* uncertainty than lab experiments? Or are they just better at sampling the population of interest, and in which case generate lower e_u and not necessarily lower $p_u(m_u)$? Interestingly, more recent papers have made use of lab-in-the-field experiments. See, for instance, Robinson (2016), Lowes et al. (2017), Lowes and Montero

(2021), Rubin and Karaja (2018), and Chaudhary et al. (2020). Does this approach lower population uncertainty $p(m_u)$, or just e_u?

Propositions 2 and 3 also has implications on comparisons of purely formal/theoretical analyses versus purely empirical tests of C_{ST}. Perhaps the most fundamental issue is whether or not any two papers are even explaining the same phenomenon to begin with. This is not a trivial matter, as social scientists can define a single social concept in many ways. The more general the concept or phenomenon — that is, one that is not limited to a particular place and point in time, the more likely it is that any two papers on the same concept are actually explaining different things. Take the case of social inequality, for instance. This is not a unique historical phenomenon, but a concept and, as such, requires some operational definition. Thus, wealth inequality is not necessarily the same as income inequality, nor racial inequality, nor gender inequality, nor status inequality. The definition and measurement of these concepts also varies across settings. It then follows that the causal chains behind these phenomena, although may be similar, cannot be exactly the same. This also means that each set of models that can respectively approximate each of the causal chains do not perfectly intersect, if at all. Caution is thus warranted when attempting to infer similar mechanisms behind even seemingly similar phenomena such as wealth and income inequality.

Even particular historical phenomena can be operationalized in many ways. For instance, what exactly is the Industrial Revolution? As discussed by Cannadine (1984) different generations of historians have meant something quite different by the term Industrial Revolution. In the 1950s and 1960s, the term industrial revolution was used to denote a dramatic acceleration of economic growth driven by rapid industrialization—"take-off" in Rostow's (1960) evocative phrase. From the 1980s onwards, however, economic historians emphasized a more gradualist perspective: economic growth and productivity growth were slow in Britain between 1770 and 1850 and real wages increased little, if at all (Crafts, 1985, e.g.). Nonetheless, there were major sectoral shifts in economic activity in this period. These shifts do not necessarily show up in the aggregate growth data. In their rehabilitation of the concept Berg and Hudson (1992, 30) note that that "aggregative studies are dogged by an inbuilt problem of identification in posing questions about the existence of an industrial revolution". More recent scholarship emphasizes the Industrial Revolution *not* because it is an exemplar of rapid growth but because it chronologically marks the onset of the modern growth era — a period when both per capita incomes and population began to rise together (see Koyama and Rubin, 2022).[7] This leaves open the question of how useful it is to apply the term industrial revolution to other cases of economic growth such as China's rapid growth since 1980.

Thus, given the same social phenomenon, some scholars can define or operationalize it as S_A, while others can define it as S_B. Note that S_A and S_B may not even belong to the same set $\{S\}$ of concurrent events/phenomena in that they can occur at different points in time. It is no wonder that the inherent models or explanations of the occurrence of the "Industrial Revolution" can differ — if $S_A \neq S_B$, then the causal chain C_{SA} behind S_A is not equivalent to the causal chain C_{SB} behind S_B. It then follows that the set of models that can approximate C_{SA} does not perfectly intersect with the set approximating C_{SB}, if there is at all an intersection.

A possibly bigger conundrum is that one cannot even know whether there is an intersection, or where the intersection of these sets of models lie. Is it then even meaningful to compare models or empirical tests across S_A and S_B? If a model or empirical exercise is better at explaining S_A than another model or empirical exercise

[7]Broadberry and Wallis (2017) view the Industrial Revolution as important only in so far as it marks a point when the rate of shrinkage declined.

explaining S_B — that is, if the uncertainty is lower in the former than in the latter, what then are we supposed to make of it, other than the researcher who generated the former is more skilled at modeling or empirical testing than the one who produced the latter? In other words, researchers studying S_A cannot debate with researchers studying S_B on causal grounds. The most they can do is debate on definitions — whether S_A or S_B is better at operationalizing, e.g. the Industrial Revolution. Even then, however, how is the debate to be judged?

Even if one were to abstract from such problems, that is, suppose one only compares two papers a and b that are about exactly the same social phenomenon S_{T1}. Given the same causal chain C_{ST}, such papers could still be using different models m_{STa} and m_{STb}. In this case, it is ambiguous whether a "good" model, i.e. which produces the lowest possible model uncertainty $\underline{m_u}$, is better or worse than a well-identified empirical exercise which generates uncertainty $\underline{c_u}$ that is based on a model with uncertainty $m_u > \underline{m_u}$. It is then possible that $\underline{c_u} \equiv \underline{o_u} \cdot p_u(m_u) \cdot \underline{e_u} > \underline{m_u}$ when m_u and $p'(m_u)$ are sufficiently large.

To provide a concrete example on how the conceptual framework proposed here can be used to assess how much one can really learn about social phenomena, the next section discusses a recent paper on social inequality.

3 Application to the Study of Status Inequality

There is a large and growing literature on social inequality, perhaps largely because the concept itself can be defined along many dimensions. Some of the recent scholarship include those that focus on one dimension, e.g. Hornbeck and Naidu (2014) on racial inequality, Grosjean and Khattar (2018) on gender inequality, Murray (2012) on class-based inequality, and Alesina et al. (2016), and Michelitch (2015) on ethnic differences. Others also consider social inequality along multiple dimensions (Case and Deaton, 2021; Carvalho and Pradelski, 2021). Below, I apply the uncertainty principle to one study of one type of social inequality, in order to demonstrate that even in a very restricted setting, uncertainties can still be non-trivial.

Desierto and Koyama (2020) provide insights on how institutions can perpetuate inequalities in social status. They take, as particular case, sumptuary laws in preindustrial Europe, which restricted the type of clothing that individuals could wear based on their social class, i.e. nobility vs. ordinary citizens. These laws proliferated in the 12th to 17th centuries, then disappeared after the 18th century. To explain this rise and fall in sumptuary legislation, the authors combine historical analytic narratives with data on the number of sumptuary laws that they compiled from multiple sources, as well as city-level and country-level panel data on various economic and political variables that have been provided by other scholars. From this rich set of information, they then propose a formal model and test some of its predictions.

In their model, the consumption of status goods – clothing in this case, by non-elites pose a status threat to ruling elites, prompting the latter to enact laws that restrict such consumption. The model then predicts a non-monotonic effect of income - sumptuary legislation initially increases with income to control the rising consumption of status goods by non-elites. However, as income continues to increase and the non-elites are able to afford even higher consumption, sumptuary laws are increasingly disregarded by non-elites, making enforcement of these laws more costly for the ruling elites. In this case, rather than enforcing the laws, ruling elites then prefer to spend more on their own status-good consumption in order to maintain their 'status distance' from the non-elites. Eventually, enforcement becomes so ineffective and the passage of additional laws decline. An additional prediction is that the initial rise in sumptuary laws is more likely when the ruling elites are less rent-seeking, since in this case they face greater status threat from non-elites.

They then empirically test the predictions. A major challenge to causal identification is that income is not exogenously given, but is likely related to many other factors that can affect ruling elites' propensity to enact sumptuary laws. To address this problem, the authors use plague outbreaks as proxy for non-elites' income, as other studies have shown that the plague increased real wages in Malthusian economies. Another challenge is cross-country heterogeneity, which the authors address by using, in addition to country-level data, city-level data from Italian city states. The drawback, however, is that this data is only available until 1500, and thus can only be used to identify the initial rise in sumptuary laws, and not the decline. To provide some evidence for the decline due to further rising income levels, the authors then use the country-level data and show (negative) correlations between income per capita and the number of sumptuary laws, and augment the analysis by qualitative accounts by historians that highlight the increasing violations of the laws and rising enforcement costs.

How much can one learn from this study? First of all, since the paper combines a model and empirical tests thereof, it likely generates less uncertainty than the same model alone. That is, $c_u < m_u$. Even if the empirical exercises are not all well-identified, i.e. even if uncertainties o_u and e_u may be large, total uncertainty $c_u = o_u \cdot p_u(m_u) \cdot e_u$ is still smaller than model uncertainty m_u alone. This is because the model is explicitly constructed, and the predictions are exactly deduced from it. Thus, assuming that $p'(m_u) \leq 1$, c_u will never be larger than m_u, even if probabilities o_u and e_u were at the possible maximum level (equal to one). Of course, it may be possible that $p'(m_u)$ is sufficiently larger than one so as to have $c_u > m_u$. However, this is unlikely for two reasons. First, because of the use of rich analytic narratives that are specific to the phenomenon of sumptuary legislation in preindustrial Europe, the uncertainty

m_u as to whether the model captures the true causal chain is unlikely to be very large. Second, precisely because there is no uncertainty behind the phenomenon in question — sumptuary laws in preindustrial Europe, there is likely to be small uncertainty as to whether the population from which the samples are drawn is the true population that experienced the phenomenon. Such population is, precisely, preindustrial Europe.

The trade off, however, is that in the precision of identifying the phenomenon, the generalization of insights is limited. In fact, an implication of Propositions 2 and 3 is that one cannot even bound the uncertainty in generalizing across other sumptuary laws (i.e. in other parts of the world and other points in time), unless there is definitive knowledge of the intersection or commonalities of such sumptuary laws with those in preindustrial Europe. Thus, the discussion of other sumptuary laws in the section on "External Validity" in Desierto and Koyama (2020), e.g. Republican Rome, Tokugawa Japan, and the Ottoman Empire, is merely an invitation to consider the possible commonalities with sumptuary laws in preindustrial Europe, and with the intersections, if any, of the causal chains that explain the laws in preindustrial Europe, and those in other contexts.

It follows that it is even harder to generalize the insights about sumptuary legislation to other laws or institutions that perpetuate status inequality, and even harder to stretch it to laws and institutions that perpetuate other kinds of social inequalities. Perhaps the only knowledge one can gain from the study, with fair amount of certainty, is on a particular institution – sumptuary legislation, that was a large part of the political economy of preindustrial Europe.

4 Concluding Remarks

What, then, of social scientists' near obsession with 'external validity' and 'generalizability'?

While there are general theories in the physical sciences, it does not follow that there are direct analogues to the social sciences. This is because a social phenomenon, to the extent that it relies on subjective definitions, is not perfectly replicable, nor are the events or circumstances leading to its occurrence. How many instances of social inequality are there, for instance, depends on what one means by social inequality. The more precise and restrictive the definition, the smaller the number of occurrences of such phenomenon. Not even a laboratory setting can replicate the phenomenon, for this would require some operationalization of the concept of social inequality. In contrast, physical phenomena can be perfectly replicable. On earth, a dropped apple always falls (unless, of course, if it is in a simulated zero-gravity environment).

Yet how much does generalizability really matter? Why can't a social phenomenon be analyzed on its own? One can argue that generalization is far more important in the physical sciences. After all, theories and laws about the physical world have enabled innovation and technological progress. But an analogous justification for establishing general theories in the social sciences — social engineering, may not be as compelling. There are arguably so many failed social policies as there are plausibly successful ones, if not more. In contrast, the light bulb is objectively a good thing. This is not to say, of course, that all technological innovation are absolutely good or useful. But those that are not often die out due to market forces. Bad social policies, however, are often sticky because political processes are inherently inefficient.

References

Alesina, Alberto, Stelios Michalopoulos and Elias Papaioannou (2016), 'Ethnic inequality', *Journal of Political Economy* **124**(2), 428–488.

Berg, Maxine and Pat Hudson (1992), 'Rehabilitating the industrial revolution.', *Economic History Review* **45**(1), 24 – 50.

Boettke, Peter J. (1997), 'Where did economics go wrong? modern economics as a flight from reality', *Critical Review* **11**(1), 11–64.

Boettke, Peter J (2000), 'Robert H. Bates, Avner Greif, Margaret Levi, Jean-Laurent Rosenthal and Barry Weingast: Analytical Narratives', *Constitutional Political*

Economy **11**(4), 377–380.
Broadberry, Stephen and John Joseph Wallis (2017), Growing, shrinking, and long run economic performance: Historical perspectives on economic development, Working Paper 23343, National Bureau of Economic Research.
URL: *http://www.nber.org/papers/w23343*
Cannadine, David (1984), 'The present and the past in the english industrial revolution 1880-1980', *Past & Present* (103), 131–172.
Carvalho, Jean-Paul and Bary Pradelski (2021), Identity and underrepresentation: Interactions between race and gender. Working paper.
Case, Anne and Angus Deaton (2021), *Deaths of Despair and the Future of Capitalism*, Princeton University Press, Princeton, NJ.
Chaudhary, Latika, Jared Rubin, Sriya Iyer and Anand Shrivastava (2020), 'Culture and colonial legacy: Evidence from public goods games', *Journal of Economic Behavior & Organization* **173**, 107–129.
Clarke, Kevin A. and David M. Primo (2011), *A Model Discipline: Political Science and the Logic of Representations*, Oxford University Press, Oxford.
Coppedge, Michael (1999), 'Thickening thin concepts and theories: Combining large N and small in comparative politics', *Comparative Politics* **31**, 465.
Coyne, Christopher J. and Peter Boettke (2015), Introduction: Austrian economics as a progressive research program, *in* C. J.Coyne and P.Boettke, eds, 'The Oxford Handbook of Austrian Economics', Oxford University Press, pp. 1–12.
Crafts, N.F.R. (1985), *British Economic Growth during the Industrial Revolution*, Clarendon Press, Oxford, U.K.
Deaton, Angus and Nancy Cartwright (2018), 'Understanding and misunderstanding randomized controlled trials', *Social Science & Medicine* **210**, 2–21.
Desierto, Desiree and Mark Koyama (2020), The Political Economy of Status Competition: Sumptuary Laws in Preindustrial Europe, CEPR Discussion Papers 14407, C.E.P.R. Discussion Papers.
URL: *https://ideas.repec.org/p/cpr/ceprdp/14407.html*
Duhem, Pierre (1906), *La Theorie Physique, Son Object, Sa Structure.*, Chevalier & Riviére, Paris. Translated as The Aim and Structure of Physical Theory. 1954. Trans. P. Wiener. Princeton University Press.
Gerring, John (2012), *Social Science Methodology: A Unified Framework*, Cambridge University Press, Cambridge.
Grosjean, Pauline and Rose Khattar (2018), 'It's Raining Men! Hallelujah? The Long-Run Consequences of Male-Biased Sex Ratios', *The Review of Economic Studies* **86**(2), 723–754.
Harding, Sandra G., ed. (1976), *Can Theories be Refuted?: Essays on the Duhem-Quine Thesis*, D. Reidel Publshing Company, Dordrecht-Holland.
Hayek, F.A. (1952), *The Counter-Revolution of Science*, University of Chicago Press, Chicago.
Hornbeck, Richard and Suresh Naidu (2014), 'When the levee breaks: Black migration and economic development in the american south', *American Economic Review* **104**(3), 963–90.
Jaki, Stanley (1974), *Science and Creation: From Eternal Cycles to an Oscillating Universe*, Scottish Academic Press., Edinburgh.
Jaki, Stanley (1978), *The Road of Science and the Ways to God*, Scottish Academic Press, Edinburgh.
Jaki, Stanley (2002), *A Mind's Matter: An Intellectual Autobiography*, Wm. B. Eerd-

mans Publishing Company, Grand Rapids.
Koyama, Mark and Jared Rubin (2022), *How We Got Rich*, Polity Press, London.
Lowes, Sara and Eduardo Montero (2021), 'The legacy of colonial medicine in central africa', *American Economic Review* **111**(4), 1284–1314.
Lowes, Sara, Nathan Nunn, James A. Robinson and Jonathan L. Weigel (2017), 'The evolution of culture and institutions: Evidence from the kuba kingdom', *Econometrica* **85**(4), 1065–1091.
Martin, Adam (2015), Austrian methodology: A review and synthesis, *in* C. J.Coyne and P.Boettke, eds, 'The Oxford Handbook of Austrian Economics', Oxford University Press, pp. 13–45.
Michelitch, Kristin (2015), 'Does electoral competition exacerbate interethnic or interpartisan economic discrimination? evidence from a field experiment in market price bargaining', *American Political Science Review* **109**(1), 43–61.
Murray, Charles (2012), *Coming Apart: The State of White America, 1960-2010*, Crown Publishing Group.
Quine, W. V. (1951), 'Main trends in recent philosophy: Two dogmas of empiricism', *The Philosophical Review* **60**(1), 20–43.
Robinson, Amanda Lea (2016), 'Nationalism and ethnic-based trust: Evidence from an african border region', *Comparative Political Studies* **49**(14), 1819–1854.
Rostow, Walt W. (1960), *The stages of economic growth : a non-communist manifesto*, Cambridge University Press, Cambridge.
Rubin, Jared and Elira Karaja (2018), The cultural transmission of trust norms: Evidence from a lab in the field on a natural experiment. Mimeo.
Ruhm, Christopher J. (2019), 'Shackling the identification police?', *Southern Economic Journal* **85**(4), 1016–1026.
Sawyer, K. R., Clive Beed and H. Sankey (1997), 'Underdetermination in economics. the Duhem-Quine thesis', *Economics and Philosophy* **13**(1), 1–23.
Skarbek, David (2020), 'Qualitative research methods for institutional analysis', *Journal of Institutional Economics* **16**(4), 409–422.
von Mises, Ludwig (1949, 1996), *Human Action*, Fox and Wilkes, San Francisco.