

NBER WORKING PAPER SERIES

THE LURE OF INCREDIBLE CERTITUDE

Charles F. Manski

Working Paper 24905

<http://www.nber.org/papers/w24905>

NATIONAL BUREAU OF ECONOMIC RESEARCH

1050 Massachusetts Avenue

Cambridge, MA 02138

August 2018

The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2018 by Charles F. Manski. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

The Lure of Incredible Certitude  
Charles F. Manski  
NBER Working Paper No. 24905  
August 2018  
JEL No. A11,D80,Z18

**ABSTRACT**

Forthright characterization of scientific uncertainty is important in principle and serves important practical purposes. Nevertheless, economists and other researchers commonly report findings with incredible certitude, reporting point predictions and estimates. To motivate expression of incredible certitude, economists often suggest that researchers respond to incentives that make the practice tempting. This temptation is the "lure" of incredible certitude. I flesh out and appraise some of the rationales that observers may have in mind when they state that incredible certitude responds to incentives. I conclude that scientific expression of incredible certitude at most has appeal in certain limited contexts. It should not be a general practice.

Charles F. Manski  
Department of Economics  
Northwestern University  
2211 Campus Drive  
Evanston, IL 60208-2600  
and NBER  
cfmanski@northwestern.edu

## 1. Introduction

### 1.1. Background

Science seeks to reveal and resolve uncertainty. Methodological studies aim to characterize scientific uncertainty. The objective is to determine the conclusions implied by specified assumptions and data.

My research on partial identification has sought to characterize a broad class of scientific uncertainties that arise when available data are used to predict population outcomes. See Manski (1995, 2003, 2007, 2013) and the articles cited therein. I have recommended that researchers first determine what can be learned when the data are combined with assumptions that are weak enough to be credible. They should then explore what further can be learned when the data are combined with stronger but less credible assumptions.

My motivations for study of partial identification have been both principled and practical. On principle, I consider forthright characterization of uncertainty to be a fundamental aspect of the scientific code of conduct. Statistical imprecision and identification problems both limit the conclusions that may be drawn in empirical research. Statistical theory characterizes the inferences that can be drawn about a study population by observing the outcomes of a sample of its members. Studies of identification characterize the inferential difficulties that persist when sample size grows without bound. Identification problems often are the dominant difficulty.

I have argued that forthright characterization of uncertainty serves important practical purposes. Viewing science as a social enterprise, I have reasoned that if scientists want people to trust what we say we know, we should be up front about what we don't know. I have suggested that inferences predicated on weak assumptions can achieve wide consensus, while ones that require strong assumptions may be subject to sharp disagreements.

I have pointed out that disregard of uncertainty when reporting research findings may harm formation of public policy. If policy makers incorrectly believe that existing analysis provides an accurate description

of history and accurate predictions of policy outcomes, they will not recognize the potential value of new research aiming to improve knowledge. Nor will they appreciate the potential usefulness of decision strategies that may help society cope with uncertainty and learn, including diversification and information acquisition.

Despite these principled and practical motivations, I have found that study of partial identification generates strong reactions. The flashpoint of controversy has been the fact that research with weak assumptions typically yields bounds on quantities of interest rather than point inferences. Some scientists are comfortable reporting findings in the form of bounds and appreciate making explicit the tradeoff between strength of assumptions and strength of findings that the bounds make plain. However, many hold firm to the traditional practice of reporting point estimates and point predictions, even though they may rest on fragile foundations or have obscure interpretations.

The traditional practice is particularly prevalent in economic policy analysis, which attempts to evaluate the impacts of past policies and predict the outcomes of potential future ones. I have repeatedly criticized policy analysis with *incredible certitude* (Manski, 2011, 2013, 2015, 2018a). Exact predictions of policy outcomes and exact estimates of the state of the economy are routine. Expressions of uncertainty are rare. I have documented that predictions and estimates often are fragile, resting on unsupported assumptions and limited data. So the expressed certitude is not credible.

To help organize thinking, Manski (2011, 2013) introduced a typology of practices that contribute to incredible certitude:

- \* conventional certitude: A prediction that is generally accepted as true but is not necessarily true.
- \* dueling certitudes: Contradictory predictions made with alternative assumptions.
- \* conflating science and advocacy: Specifying assumptions to generate a predetermined conclusion.
- \* wishful extrapolation: Using untenable assumptions to extrapolate.
- \* illogical certitude: Drawing an unfounded conclusion based on logical errors.
- \* media overreach: Premature or exaggerated public reporting of policy analysis.

The definitions of conventional and dueling certitudes refer to predictions, but they apply as well to estimates of realized quantities. I have provided illustrative examples and offered suggestions to improve practices.

How do researchers motivate reporting findings with incredible certitude? Economists often suggest that researchers respond to incentives. Analysis supporting this suggestion has been lacking, but I have witnessed many offhand remarks. A common perception among economists who act as consultants to firms or policy makers is that the public is either unwilling or unable to cope with uncertainty. Hence, they argue that pragmatism dictates provision of point predictions and estimates, even though they may not be credible.

To cite some examples, over fifty years ago, Morgenstern (1963) remarked that federal statistical agencies may perceive a political incentive to express incredible certitude about the state of the economy when they publish official economic statistics. He wrote (p. 11):

"All offices must try to impress the public with the quality of their work. Should too many doubts be raised, financial support from Congress or other sources may not be forthcoming. More than once has it happened that Congressional appropriations were endangered when it was suspected that government statistics might not be 100 percent accurate. It is natural, therefore, that various offices will defend the quality of their work even to an unreasonable degree."

The econometrician Jerry Hausman put it this way at a conference in 1988, when I presented in public some of my early findings on partial identification: "You can't give the client a bound. The client needs a point." Douglas Holtz-Eakin, former director of the Congressional Budget Office (CBO), told me in 2010 that he expected Congress would be highly displeased if the CBO were to express uncertainty in the official predictions that it makes for the future impact on the federal debt of pending legislation.

In the mid-1990s, I summarized multiple conversations by writing (Manski, 1995, p. 3):

"The scientific community rewards those who produce strong novel findings. The public, impatient for solutions to its pressing concerns, rewards those who offer simple analyses leading to unequivocal policy recommendations. These incentives make it tempting for researchers to maintain assumptions far stronger than they can persuasively defend, in order to draw strong conclusions."

For short, I now call this temptation the "lure" of incredible certitude.

Incredible certitude is not peculiar to economics. I have observed many manifestations throughout the social sciences and in medical research; see Manski (2013, 2018b). Yet some fields endeavor to be forthright about uncertainty.

I particularly have in mind climate science, which has sought to predict how greenhouse gas emissions affect the trajectory of atmospheric temperature and sea level. Published articles on climate science often make considerable effort to quantify uncertainty. See, for example, Knutty *et al.* (2010), McGuffie and Henderson-Sellers (2005), McWilliams (2007), Parker (2006, 2013), Palmer *et al.* (2005), and Stainforth *et al.* (2007). The attention paid to uncertainty in the periodic reports of the Intergovernmental Panel on Climate Change (IPCC) is especially notable; see Mastrandrea *et al.* (2010).

## 1.2. This Paper

The principled and practical arguments against research with incredible certitude are strong. Nevertheless, such research remains common in economics and elsewhere. I am unaware of systematic enquiries that justify expression of incredible certitude. I am aware only of casual references to incentives such as those that I quoted above.

This paper strives to flesh out and appraise some of the rationales that observers may have in mind when they state that incredible certitude responds to incentives. I open the question and make some progress. I do not claim to fully settle the matter.

The first task is to document the prevalence of incredible certitude. I have previously done so to a considerable extent, with focus on economic policy analysis (Manski, 2011, 2013), government release of official economic statistics (Manski, 2015), and evidence-based medicine (Manski, 2018b). After an opening discussion of certitude in faith and philosophy, Section 2 distinguishes two qualitatively different scientific practices and provides illustrative cases of each.

One practice combines available data with questionable assumptions to form a point prediction or estimate of a quantity of substantive interest. I discuss CBO scoring of legislation, official estimates of GDP growth and the household income distribution, and provision by health agencies of risk-assessment tools that predict the future onset and outcomes of diseases. I also discuss cases of dueling certitudes, where different studies report mutually contradictory findings, each reported with certitude. Here I use criminal justice research to illustrate.

The other practice acknowledges that one cannot form a credible point prediction or estimate of a quantity of substantive interest. Rather than express uncertainty about this quantity, researchers change the objective and report a point prediction or estimate of another quantity that is not of substantive interest. Thus, they sacrifice relevance for certitude. Sacrificing relevance for certitude does not imply incredible certitude per se, but it does when authors or readers misinterpret the quantities being reported. I use medical research to illustrate, discussing use of odds ratios to measure health risks, the prevailing focus on internal validity in research reporting randomized trials, and meta-analysis of disparate studies.

Section 3 poses and assesses several possible rationales for assertions that incredible certitude responds to incentives. Each rationale presumes that incredible certitude aims to enhance social welfare. The question in each case is the strength of the foundation for thinking that this aim is achieved.

I first discuss a psychological argument asserting that scientific expression of incredible certitude is necessary because the public is unable to cope with uncertainty. I conclude that this argument has a weak empirical foundation. Research may support the claim that some persons are intolerant of some types of uncertainty, but it does not support the claim that this is a general problem of humanity. The reality appears to be that humans are quite heterogeneous in the ways that they deal with uncertainty.

I next discuss a bounded-rationality argument asserting that incredible certitude may be useful as a device to simplify decision making. I consider the usual formalization of decision under uncertainty in which a decision maker perceives a set of feasible states of nature and must choose an action without knowledge

of the actual state. Suppose that evaluation of actions requires effort. Then it simplifies decision making to restrict attention to one state of nature and optimize as if this is truth, rather than make a choice that acknowledges uncertainty. However, the result may be degradation of decision making if the presumed certitude is not credible.

A third rationale arises from consideration of collective decision making. The argument is that social acceptance of conventional certitudes may be a useful coordinating device, preventing coordination failures that may occur if persons deal with uncertainty in different ways. This rationale is broadly similar to the bounded-rationality one. Both assert that incredible certitude simplifies decision making, individual or collective as the case may be.

I conclude in Section 4 that scientific expression of incredible certitude at most has practical appeal in certain limited contexts. On principle, characterization of uncertainty is fundamental to science. Hence, researchers should generally strive to convey uncertainty clearly.

## 2. Manifestations of Incredible Certitude

### 2.1. Certitude in Faith and Philosophy

#### 2.1.1. Religious Dogma

While my concern is incredible certitude in modern science, it is worth keeping in mind that expression of uncertainty is an ancient human issue. Religious dogma provides extreme manifestations of incredible certitude. Hebrew prayers asserting the existence and power of God end with the congregation stating "Amen," which is variously interpreted to mean "certainty," "truth," or "I believe." The Apostles' Creed of Christianity asserts that the speaker believes in basic tenets of the faith and concludes "I believe in the Holy



Spirit, the holy catholic Church, the communion of saints, the forgiveness of sins, the resurrection of the body, and the life everlasting. Amen." No proof of these tenets is given and no space is left for uncertainty. The faith asks that one simply believe.

Religious dogma is a conventional certitude in a society with a consensus faith. Dueling certitudes occur when persons hold different faiths whose dogmas are inconsistent with one another. It is sometimes said that dueling certitudes may be useful as a device to promote learning. The ancient idea of dialectic proposes that debating contradictory perspectives can be an effective way to determine truth. However, history presents numerous examples of bitter conflicts that result from dueling religious certitudes.

### 2.1.2. Occam's Razor

Classical and enlightenment philosophers manifest a spectrum of views about uncertainty. Some assert that they know basic truths while others express uncertainty. I will focus on one persistent idea in the philosophy of science, namely that a scientist should choose one hypothesis among those that are consistent with the available data.

Researchers often refer to *Occam's Razor*, the medieval philosophical declaration that "Plurality should not be posited without necessity." An article by Duignan (2017) in the *Encyclopaedia Britannica* gives the usual modern interpretation of this cryptic statement, remarking that: "The principle gives precedence to simplicity; of two competing theories, the simplest explanation of an entity is to be preferred." The philosopher Richard Swinburne writes (Swinburne, 1997, p. 1):

"I seek...to show that—other things being equal—the simplest hypothesis proposed as an explanation of phenomena is more likely to be the true one than is any other available hypothesis, that its predictions are more likely to be true than those of any other available hypothesis, and that it is an ultimate a priori epistemic principle that simplicity is evidence for truth."

The choice criterion offered here is as imprecise as the one given by Occam. What do Duignan and Swinburne mean by "simplicity?"

Among economists, Milton Friedman expressed the Occam perspective in an influential methodological essay. Friedman (1953) placed prediction as the central objective of science, writing (p. 5): “The ultimate goal of a positive science is the development of a ‘theory’ or ‘hypothesis’ that yields valid and meaningful (i.e. not truistic) predictions about phenomena not yet observed.” He went on to say (p. 10):

“The choice among alternative hypotheses equally consistent with the available evidence must to some extent be arbitrary, though there is general agreement that relevant considerations are suggested by the criteria ‘simplicity’ and ‘fruitfulness,’ themselves notions that defy completely objective specification.”

Thus, Friedman counseled scientists to choose one hypothesis, even though this may require the use of “to some extent . . . arbitrary” criteria. He did not explain why scientists should choose one hypothesis from many. He did not entertain the idea that scientists might offer predictions under the range of plausible hypotheses that are consistent with the available evidence.

However one may operationalize philosophical dicta for choosing a single hypothesis, the relevance of philosophical thinking to economics is not evident. In economic analysis, knowledge is instrumental to the objective of making good decisions. When philosophers discuss the logical foundations and human construction of knowledge, they do so without posing this objective. Does use of a criterion such as “simplicity” to choose one hypothesis promote good decision making? As far as I am aware, philosophers have not addressed this essential economic question.

## 2.2. Conventional Certitude in Government Predictions and Estimates

Recall that a conventional certitude is a prediction or estimate that is generally accepted as true but that is not necessarily true. Official government prediction and estimation practices provide notable illustrations. Official statistics are usually reported as point predictions or estimates. Lacking guidance, some users of the statistics may naively assume that they are accurate. Persons who understand that the statistics are subject

to error must fend for themselves and conjecture the error magnitudes. Thus, users may misinterpret the information that the statistics provide.

I begin with a leading economic case of prediction and then describe two leading cases of estimation. To illustrate in another domain, I also discuss health risk assessment.

### 2.2.1. CBO Scoring

Conventional certitude is exemplified by Congressional Budget Office scoring of federal legislation. The CBO was established in the Congressional Budget Act of 1974. The Act has been interpreted as mandating the CBO to provide point predictions (*scores*) of the budgetary impact of legislation. CBO scores are conveyed in letters that the Director writes to leaders of Congress. They are unaccompanied by measures of uncertainty, even though the budgetary impacts of complex changes to federal law are difficult to foresee.

Credible scoring is particularly difficult to achieve when proposed legislation may significantly affect the behavior of individuals and firms, by changing the incentives they face to work, hire, make purchases, and so on. Serious policy analysts recognize that scores for complex legislation are fragile numbers, derived from numerous untenable assumptions. Considering the related matter of scoring the effects of tax changes on federal revenues, Auerbach (1996) wrote (p. 156): “in many instances, the uncertainty is so great that one honestly could report a number either twice or half the size of the estimate actually reported.” CBO scores exemplify conventional certitude because they have achieved broad acceptance. They are used by both Democratic and Republican members of Congress. Media reports largely take them at face value.

Well-known examples of scoring complex legislation are the scoring of the Patient Protection and Affordable Care Act of 2010, commonly known as Obamacare or the ACA, and of the American Health Care Act of 2017, which sought to partially repeal the ACA. In March 2010 the CBO and the Joint Committee on Taxation (JCT) jointly scored the combined consequences of the ACA and the Reconciliation Act of 2010 and reported (Elmendorf, 2010, p.2): “enacting both pieces of legislation . . . would produce a net reduction

of changes in federal deficits of \$138 billion over the 2010-2019 period as a result of changes in direct spending and revenue.” Media reports largely accepted the CBO score as fact without questioning its validity, the hallmark of conventional certitude.

In March 2017 the CBO and JCT scored the American Health Care Act and reported (Congressional Budget Office, 2017, p. 1): "enacting the legislation would reduce federal deficits by \$337 billion over the 2017-2026 period." The CBO verbally acknowledged uncertainty in this prediction, writing (p. 3-4):

"The ways in which federal agencies, states, insurers, employers, individuals, doctors, hospitals, and other affected parties would respond to the changes made by the legislation are all difficult to predict, so the estimates in this report are uncertain. But CBO and JCT have endeavored to develop estimates that are in the middle of the distribution of potential outcomes."

However, the point prediction of a deficit reduction of \$337 billion was not accompanied by quantitative measurement of uncertainty.

The CBO has established a reputation for impartiality. Perhaps it is best to have the CBO express certitude when it scores legislation, even if the certitude is conventional rather than credible. But I have worried that the social contract to take CBO scores at face value will break down. I have suggested that it would be better for the CBO to protect its reputation than to have some disgruntled group in the government or the media declare that the emperor has no clothes (Manski, 2011, 2013).

A simple approach would be to provide interval forecasts of the budgetary impacts of legislation. The CBO would produce two scores for a bill, a low and a high score, and report both. For example, the CBO could report the 0.10 and 0.90 quantiles of the distribution of potential outcomes that it referenced when scoring the American Health Care Act of 2017. Or it could present a full probabilistic forecast in a graphical fan chart such as the Bank of England uses to predict GDP growth (see the discussion in Section 2.2.2). If the CBO must provide a point prediction, it can continue to do so, with some convention used to locate the point within the interval forecast.

### 2.2.2. National Income Accounts

Further leading examples of conventional certitude are the official economic statistics published by federal statistical agencies including the Bureau of Economic Analysis, Bureau of Labor Statistics, and Census Bureau. These agencies respectively report point estimates of GDP growth, unemployment, and household income. Agency staff know that official statistics suffer from sampling and non-sampling errors. Yet the practice has been to report statistics with only occasional measurement of sampling errors and no measurement of non-sampling errors. The media and the public generally accept the estimates as reported, making them instances of conventional certitude.

Considering the uncertainties in official statistics, I have found it useful to refine the general problem of conventional certitude, distinguishing errors in measurement of well-defined concepts from uncertainty about the concepts themselves. I have also found it useful to distinguish transitory and permanent measurement problems. Thus, Manski (2015) separately discussed transitory statistical uncertainty, permanent statistical uncertainty, and conceptual uncertainty. I give a notable example of transitory uncertainty here and of permanent uncertainty in Section 2.2.3. Manski (2015) discusses seasonal adjustment of time series as a problem of conceptual uncertainty.

Transitory statistical uncertainty arises because data collection takes time. Agencies may release a preliminary statistic with incomplete data and revise as new data arrives. Uncertainty diminishes as data accumulates. A leading example is Bureau of Economic Analysis (BEA) initial measurement of GDP and revision of the estimate as new data arrives. The BEA reports multiple vintages of quarterly GDP estimates. An 'advance' estimate combines data available one month after the end of a quarter with trend extrapolations. 'Second' and 'third' estimates are released after two and three months, when new data become available. A 'first annual' estimate is released in the summer, using data collected annually. There are subsequent annual and five-year revisions. Yet the BEA reports GDP estimates without quantitative measures of uncertainty.

A publication by BEA staff explains the practice of reporting estimates without measures of error as a response to the presumed wishes of the users of GDP statistics. Fixler, Greenaway-McGrevy, and Grimm (2014) state (p. 2):

"Given that BEA routinely revises its estimates during the course of a year, one might ask why BEA produces point estimates of GDP instead of interval estimates. . . Although interval estimates would inform users of the uncertainty surrounding the estimates, most users prefer point estimates, and so they are featured."

BEA analysts have provided an upbeat perspective on the accuracy of GDP statistics; see Fixler, Greenaway-McGrevy, and Grimm (2011). But Croushore (2011) offers a more cautionary perspective, writing (p. 73):

"Until recently, macroeconomists assumed that data revisions were small and random and thus had no effect on structural modeling, policy analysis, or forecasting. But realtime research has shown that this assumption is false and that data revisions matter in many unexpected ways."

Communication of the transitory uncertainty of GDP estimates should be relatively easy to accomplish. The historical record of revisions has been made accessible for study in two 'realtime' data sets maintained by the Philadelphia and St. Louis Federal Reserve Banks; see Croushore (2011) for definition of 'realtime.' Measurement of transitory uncertainty in GDP estimates is straightforward if one finds it credible to assume that the revision process is time-stationary. Then historical estimates of the magnitudes of revisions can credibly be extrapolated to measure the uncertainty of future revisions.

The BEA could communicate uncertainty as a probability distribution via a fan chart, as the Bank of England does regularly. See Aikman *et al.* (2011) for commentary on the thinking underlying the Bank's use of fan charts to communicate uncertainty.

### 2.2.3. Household Income and Unemployment Statistics

Permanent statistical uncertainty arises from incompleteness or inadequacy of data collection that is not resolved over time. Sources include sampling error due to finite sample size and non-sampling error due to

nonresponse and misreporting. I focus here on nonresponse to employment and income questions in the Current Population Survey.

Each year the U. S. Census Bureau reports statistics on the household income distribution based on data collected in a supplement to the CPS. The Census Bureau's annual *Current Population Report* provides statistics characterizing the income distribution and measures sampling error by providing 90-percent confidence intervals for various estimates. The report does not measure non-sampling errors. A supplementary document describes some sources of non-sampling error, but it does not quantify them.

Each month, the BLS issues a news release reporting a point estimate of the unemployment rate for the previous month, based on data collected in the monthly CPS. A Technical Note issued with the release contains a section on *Reliability of the estimates* that acknowledges the possibility of errors (Bureau of Labor Statistics, 2018). The Note describes the use of standard errors and confidence intervals to measure sampling error. It states that non-sampling errors

"can occur for many reasons, including the failure to sample a segment of the population, inability to obtain information for all respondents in the sample, inability or unwillingness of respondents to provide correct information on a timely basis, mistakes made by respondents, and errors made in the collection or processing of the data."

The Note does not measure the magnitudes of non-sampling errors.

When the Census Bureau and BLS report point estimates of statistics on household income and employment, they assume that nonresponse is random conditional on specified observed covariates of sample members. This assumption, which implies the absence of non-sampling error, is implemented as weights for unit nonresponse and imputations for item nonresponse. CPS documentation of its imputation approach offers no evidence that the method yields a distribution for missing data that is close to the actual distribution. Another Census document describing the American Housing Survey is revealing. It states (U.S. Census Bureau, 2011): "The Census Bureau does not know how close the imputed values are to the actual values." Indeed, lack of knowledge of the closeness of imputed values to actual ones is common.

Research on partial identification shows how to measure the potential consequences of non-sampling error due to nonresponse without making assumptions about the nature of the missing data. To begin, one contemplates all values that the missing data can take. Then the data yield interval estimates of official statistics that make no assumptions about the values of missing data. The literature derives intervals for population means and quantiles. The intervals have simple forms, the lower and upper bounds being the values that the estimate would take if all missing data were to take the smallest or largest logically possible value. The literature shows how to form confidence intervals that jointly measure sampling and nonresponse error. See Manski (1989, 1994, 2003), Horowitz and Manski (1998), and Imbens and Manski (2004) for original research articles and Manski (2007) for a textbook exposition. Manski (2016) gives an application to CPS data on household income.

Interval estimates of official statistics that place no assumptions on the values of missing data are easy to understand and simple to compute. One might therefore think that it would be standard practice for government statistical agencies to report them, but official statistics are not reported this way. It is sometimes said that such interval estimates are "too wide to be informative," but I recommend that statistical agencies report them. Wide bounds reflect real data uncertainties that cannot be washed away by assumptions lacking credibility.

The above does not imply that statistical agencies should refrain from making assumptions about nonresponse. Interval estimates making no assumptions may be excessively conservative if agency analysts have some understanding of the nature of nonresponse. There is much middle ground between interval estimation with no assumptions and point estimation assuming that nonresponse is conditionally random. The middle ground obtains interval estimates using assumptions that may include random nonresponse as one among various possibilities. Manski (2016) poses some alternatives that agencies may want to consider.



#### 2.2.4. The Breast Cancer Risk Assessment Tool

It has become common for health agencies to provide online tools that predict the probability that a person with specified attributes will develop some disease or experience certain health outcomes. These prediction tools report probabilistic rather than deterministic predictions, so one might view them as expressing uncertainty adequately. However, the tools report precise probabilities when the available information on health risks does not support doing so. Hence, the tools manifest incredible certitude, in the sense of non credible certainty about the correct value of the probability.

A prominent case is the Breast Cancer Risk Assessment (BCRA) Tool of the National Cancer Institute (2011). The BCRA Tool reports the probability that a woman will develop breast cancer conditional on eight attributes characterizing her family and health history. The Tool has become widely used in clinical practice and is an important input to the clinical practice guidelines differentiating women who should receive routine breast cancer screening from those who warrant prophylactic treatment. For example, National Comprehensive Cancer Network (2017) recommends routine screening if the predicted probability of breast cancer in the next five years is below 0.017 and prophylactic treatment if the probability is higher.

A user of the BCRA Tool who inputs the required personal attributes receives in response a precise probability of disease development. Yet statistical imprecision and identification problems make these risk assessments uncertain. Without discussion of uncertainty, clinicians and patients may mistakenly believe that precise probabilistic risk assessments are accurate. If uncertainty is not quantified, those who recognize the presence of uncertainty cannot evaluate the degree to which assessments may be inaccurate. I call attention here to statistical imprecision in the predictions.

The BCRA Tool implements a modified form of the Gail Model introduced in the research article of Gail *et al.* (1989). The article is careful to call attention to statistical imprecision in its estimates of the probability of developing breast cancer over various future age intervals. It describes a general procedure

for estimating confidence intervals for its risk assessments. It reports illustrative computations of 95% confidence intervals for two women with different specified attributes.

The computed confidence intervals are revealing. They vary considerably in width, indicating that statistical imprecision is much more an issue when assessing risks for some women than for others. While the Gail *et al.* article is forthright in its evaluation of statistical imprecision, the BCRA Tool that implements a version of the Gail Model does not report confidence intervals. Indeed, the website that houses the BCRA Tool makes no mention of statistical imprecision.

### 2.3. Dueling Certitudes in Criminal Justice Research

Dueling certitudes—contradictory predictions made with alternative assumptions—are common in research on controversial policy questions. Research on criminal justice policy provides many illustrations. I discuss two here.

#### 2.3.1. The RAND and IDA Studies of Cocaine Control Policy

During the mid-1990s, two studies of cocaine control policy played prominent roles in discussions of federal policy towards illegal drugs. One was performed by analysts at RAND (Rydell and Everingham, 1994) and the other by analysts at the Institute for Defense Analyses (IDA) (Crane, Rivolo, and Comfort, 1997). The two studies posed similar hypothetical objectives for cocaine-control policy, namely reduction in cocaine consumption in the United States by one percent. Both studies predicted the cost of using certain policies to achieve this objective. However, RAND and IDA used different assumptions and data to reach dramatically different policy conclusions.

The authors of the RAND study specified a model of the supply and demand for cocaine that aimed to characterize the interaction of producers and users and the process through which alternative cocaine-control

policies may affect consumption and prices. They used this model to evaluate various demand-control and supply-control policies and concluded that drug treatment, a demand-control policy, is much more effective than any supply policy. The authors of the IDA study examined the time-series association between source-zone interdiction activities and retail cocaine prices. They concluded that source-zone interdiction, a supply-control policy, is at least as effective as is drug treatment.

When they appeared, the RAND and IDA studies drew attention to the ongoing struggle over federal funding of drug control activities. The RAND study was used to argue that funding should be shifted towards drug treatment programs and away from activities to reduce drug production or to interdict drug shipments. The IDA study, undertaken in part as a re-analysis of the RAND findings, was used to argue that interdiction activities should be funded at present levels or higher.

At the request of the Office of National Drug Control Policy (ONDCP), the National Research Council Committee on Data and Research for Policy on Illegal Drugs assessed the RAND and IDA studies; see National Research Council (1999). After examining the two studies, the Committee concluded that neither constitutes a persuasive basis for the formation of cocaine control policy. The Committee concluded that neither the RAND nor the IDA study provides a credible estimate of what it would cost to use alternative policies to reduce cocaine consumption in the United States.

### 2.3.2. How Do Right-to-Carry Laws Affect Crime Rates?

A considerable body of research on crime in the United States has used data on county or state crime rates to evaluate the impact of laws allowing individuals to carry concealed handguns—so called right-to-carry (RTC) laws. Theory alone cannot predict even the direction of the impact. The knowledge or belief that potential victims may be carrying weapons may deter commission of some crimes but may escalate the severity of criminal encounters. Ultimately, how allowing individuals to carry concealed weapons affects crime is an empirical question.

Lott (2010) describes some of this empirical research in a book with the provocative and unambiguous title *More Guns, Less Crime*. Yet, despite dozens of studies, the full body of research provides no clear insight on whether “more guns” leads to less crime. Some studies find that RTC laws reduce crime, others find that the effects are negligible, and still others find that such laws increase crime. In a series of papers starting in 1997, Lott and coauthors have argued forcefully that RTC laws have important deterrent effects which can play a role in reducing violent crime. Lott and Mustard (1997) and Lott (2010), for example, found that RTC laws reduce crime rates in every violent crime category by between 5 and 8 percent. Using different models and revised/updated data, however, other researchers have found that RTC laws either have little impact or may increase violent crime rates. See, for example, Black and Nagin (1998), Duggan (2001), Aneja *et al.* (2011), and Durlauf *et al.* (2016).

This sharp disagreement may seem surprising. How can researchers using similar data draw such different conclusions? In fact, it has long been known that inferring the magnitude and direction of treatment effects is an inherently difficult undertaking. Suppose that one wants to learn how crime rates (an outcome of interest) would differ with and without a RTC law (a treatment) in a given place and time. Data cannot reveal counterfactual outcomes. That is, data cannot reveal what the crime rate in a RTC state would have been if the state had not enacted the law. Nor can data reveal what the crime rate in a non-RTC state would have been if a RTC law had been in effect. To identify the law’s effect, one must somehow “fill in” the missing counterfactual observations. This requires making assumptions that cannot be tested empirically. Different assumptions may yield different inferences, hence dueling certitudes.

Empirical research on RTC laws has struggled to find consensus on a set of credible assumptions. Reviewing the literature, the National Research Council Committee to Improve Research Information and Data on Firearms concluded that it is not possible to infer a credible causal link between RTC laws and crime using the current evidence (National Research Council, 2005). Indeed, the Committee concluded that (p. 150): “additional analysis along the lines of the current literature is unlikely to yield results that will

persuasively demonstrate” this link. The Committee found that findings are highly sensitive to model specification. Yet there is no solid foundation for particular assumptions and, as a result, no obvious way to prefer specific results. Hence, drawing credible precise findings that lead to consensus about the impact of RTC laws has been impossible.

The antidote to dueling certitudes about the effect on crime of RTC laws is to recognize uncertainty by generating a set of estimates under alternative assumptions. To formalize this idea in a flexible manner, Manski and Pepper (2018) studies the conclusions implied by relatively weak 'bounded variation' assumptions that restrict variation in treatment response across places and time. The results are findings that bound the crime effect of RTC laws. Considering a set of alternative assumptions makes transparent how assumptions shape inference.

#### 2.4. Sacrificing Relevance for Certitude in Medical Research

Researchers often are aware that they cannot form a credible point prediction or estimate of a quantity of interest. They could face up to uncertainty and determine what they can credibly infer about the quantity, perhaps obtaining a bound. However, the lure of incredible certitude being strong, they often respond differently. They change the objective and focus on another quantity that is not of substantive interest but that can be predicted or estimated credibly. Thus, they sacrifice relevance for certitude.

Notable scientists have critiqued this common practice. The statistician John Tukey wrote (Tukey, 1962, p.13-14): "Far better an approximate answer to the right question, which is often vague, than an exact answer to the wrong question, which can always be made precise." Many cite some version of the joke about the drunk and the lamppost. Noam Chomsky has been quoted as putting it this way (Barsky, 1998, p. 95): "Science is a bit like the joke about the drunk who is looking under a lamppost for a key that he has lost on the other side of the street, because that's where the light is." My econometrics colleague Arthur Goldberger

was fond of quoting a line from a song by Stephen Stills, who attributed it to the musician Billy Preston: "If you can't be with the one you love, love the one you're with."

Sacrificing relevance for certitude does not imply incredible certitude if everyone understands that the quantity being estimated or predicted is not of substantive interest. The problem is that authors may not be forthright about this or readers may misinterpret findings. I provide three illustrations, focusing on medical research.

#### 2.4.1. The Odds Ratio and Public Health

In a well-known text on epidemiology, Fleiss (1981, p. 92) states that retrospective studies of disease do not yield policy-relevant predictions and so are “necessarily useless from the point of view of public health.” Nevertheless, he goes on to say that “retrospective studies are eminently valid from the more general point of view of the advancement of knowledge.” What Fleiss means in the first statement is that retrospective studies do not provide data that enable credible point estimation of *attributable risk*, a quantity of substantive interest in public health. The second statement means that retrospective studies enable credible point estimation of the *odds ratio*, a quantity that is not of substantive interest but that is widely reported in epidemiological research. I explain here, drawing on Manski (2007, Chapter 5).

The term “retrospective studies” refers to a sampling process that is also known to epidemiologists as *case-control* sampling and to econometricians studying behavior as *choice-based* sampling (Manski and Lerman, 1977). I call it *response-based* sampling here, as in Manski (2007). Formally, consider a population each of whose members is described by covariates  $x$  and a response (or outcome)  $y$ . Consider inference on the response probabilities  $P(y|x)$  when the population is divided into response strata and random samples are drawn from each stratum. This is response-based sampling.

In a simple case prevalent in epidemiology,  $y$  is a binary health outcome and  $x$  is a binary risk factor. Thus,  $y = 1$  if a person becomes ill and  $y = 0$  otherwise, while  $x = 1$  if the person has the risk factor and  $x$

= 0 otherwise. In a classic example,  $y$  denotes the presence of lung cancer and  $x$  denotes whether a person is a smoker. Response-based sampling draws random samples of ill and healthy persons. This reveals the distributions of the risk factor among those who are ill and healthy; that is,  $P(x|y = 1)$  and  $P(x|y = 0)$ . It does not reveal  $P(y|x)$ .

A basic concern of research in public health is to learn how the probability of illness varies across persons who do and who do not have a risk factor. Attributable risk is the difference in illness probability between these groups; that is,  $P(y = 1|x = 1) - P(y = 1|x = 0)$ . Another measure of the variation of illness with the risk factor is the ratio  $P(y = 1|x = 1)/P(y = 1|x = 0)$ , called *relative risk*.

Texts on epidemiology discuss both relative and attributable risk, but empirical research has focused on relative risk. This focus is hard to justify from the perspective of public health. The health impact of a risk factor presumably depends on the number of illnesses affected; that is, on attributable risk times the size of the population. The relative risk statistic is uninformative about this quantity.

For example, consider two scenarios. In one, the probability of lung cancer conditional on smoking is 0.12 and conditional on nonsmoking is 0.08. In the other, these probabilities are 0.00012 and 0.00008. The relative risk in both scenarios is 1.5. Attributable risk is 0.04 in the first scenario and 0.00004 in the second. The first scenario is clearly much more concerning to public health than the second. The relative risk statistic does not differentiate the scenarios, but attributable risk does.

Given that attributable risk is more relevant to public health, it seems odd that epidemiological research has emphasized relative risk rather than attributable risk. Indeed, the practice has long been criticized; see Berkson (1958), Fleiss (1981, Section 6.3), and Hsieh, Manski, and McFadden (1985). The rationale, such as it is, rests on the widespread use in epidemiology of response-based sampling.

The data generated by response-based sampling do not point-identify attributable risk. Fleiss (1981, p. 92) remarked that “retrospective studies are incapable of providing estimates” of attributable risk. Manski (2007) proves that these data do yield a bound.

Cornfield (1951) showed that the data from response-based sampling point-identify the odds ratio, defined as  $[P(y = 1 | x = 1)/P(y = 0 | x = 1)]/[P(y = 1 | x = 0)/P(y = 0 | x = 0)]$ . He also observed that when  $P(y = 1)$  is close to zero, a condition called the "rare-disease" assumption, the odds ratio approximately equals relative risk. The rare-disease assumption is credible when considering some diseases. In such cases, epidemiologists have used the odds ratio as a point estimate of relative risk.

Cornfield's finding motivates the widespread epidemiological practice of using response-based samples to estimate the odds ratio and then invoking the rare-disease assumption to interpret the odds ratio as relative risk. Fleiss's statement that retrospective studies are "valid from the more general point of view of the advancement of knowledge" endorses this practice. Thus, use of the odds ratio to point-estimate relative risk sacrifices relevance for certitude.

#### 2.4.2. Randomized Trials and the Primacy of Internal Validity

Randomized trials of treatment response have long enjoyed a favored status in medical research and have increasingly acquired this status in the social sciences. However, the treatment response studied in a trial may differ considerably from the response that a clinician or policy maker would find of substantive interest. Focusing on medical trials, Manski (2018b) documents three reasons.

First, the study populations enrolled in trials often differ from the patient populations that clinicians treat. Participants in trials are volunteers and are typically restricted to persons who lack co-morbidities. Second, the treatments assigned in trials often differ from those that would be assigned in clinical practice. Trial participants may receive more intensive care than they would in practice and drug treatments generally are blinded. Third, researchers performing trials often measure surrogate outcomes rather than health outcomes that really matter for patient care. For example, trials studying cardiovascular disease may measure blood pressure and cholesterol levels rather than the occurrence of heart attacks and strokes. For these and



other reasons, the point estimates of treatment effects commonly reported in articles on trials often are not credible estimates of treatment effects of substantive interest.

Seeking to justify the point estimates obtained in trials, researchers in public health and the social sciences often cite Donald Campbell, who distinguished between the internal and external validity of studies of treatment response (Campbell and Stanley, 1963; Campbell, 1984). A study is said to have *internal validity* if its findings for the study population are credible. It has *external validity* if one finds it credible to extrapolate the findings to a setting of substantive interest. In this terminology, the appeal of randomized trials is their internal validity.

Campbell argued that studies of treatment response should be judged first by their internal validity and secondarily by their external validity. In practice, researchers commonly neglect external validity. Analyses of trials focus on the outcomes measured with the treatments assigned in the study population. Research articles may offer verbal conjectures on external validity in the discussion sections of their papers, but they do not assess external validity quantitatively. Thus, relevance is sacrificed for certitude.

The doctrine of the primacy of internal validity has been extended from randomized trials to observational studies. When considering the design and analysis of observational studies, Campbell and his collaborators have recommended that researchers aim to emulate as closely as possible the conditions of a randomized experiment, even if this requires focus on a study population that differs materially from the population of interest. The statistician Paul Rosenbaum put it this way (Rosenbaum, 1999, p. 263):

“In a well-conducted laboratory experiment one of the rarest of things happens: The effects caused by treatments are seen with clarity. Observational studies of the effects of treatments on human populations lack this level of control but the goal is the same. Broad theories are examined in narrow, focused, controlled circumstances.”

Among economists, this perspective on observational studies has been championed by those who advocate study of a *local average treatment effect* (LATE). This is defined as the average treatment effect within the sub-population of persons whose received treatment would be modified by altering the value of

an instrumental variable; see Imbens and Angrist (1994) and Angrist, Imbens, and Rubin (1996). Local average treatment effects generally are not quantities of substantive interest; see Manski (1996, 2007), Deaton (2009), and Heckman and Urzua (2009). Their study has been motivated by the fact that they are point-identified given certain assumptions that are sometimes thought credible. See Imbens (2010) for a defense of study of LATEs with the provocative title "Better LATE than Nothing."

#### 2.4.3. Meta-Analysis of Disparate Studies

Difficulties arise when researchers attempt to combine findings from multiple studies. It is easy to understand the impetus for combination of findings. Readers want to interpret the mass of information in empirical research. The question is how to interpret this information sensibly.

Statisticians have proposed *meta-analysis* in an effort to provide an objective methodology for combining the findings of multiple studies. Meta-analysis was originally developed to address a purely statistical problem. Suppose that multiple trials or observational studies have been performed on the same study population, each drawing an independent random sample. The most precise way to use the data combines them into one sample. Suppose that the raw data are unavailable. Instead, multiple parameter estimates are available, each computed with the same method using data from a different sample. Meta-analysis proposes methods to combine the estimates. The usual proposal is to compute a weighted-average of the estimates, the weights varying with sample size.

The original concept of meta-analysis is uncontroversial, but it has limited applicability. It is rare that multiple independent studies are performed on the same population. It is more common for multiple studies to be performed on distinct populations that may have different distributions of treatment response. The protocols for administration of treatments and the measurement of outcomes may vary across studies as well.

Meta-analyses are performed often in such settings, computing weighted averages of estimates for distinct study populations and study designs. Averages computed with subjective weights are called *Bayesian weighted averages* or *Bayesian model averaging*.

The problem is that it may not be clear how to define and interpret a weighted average of the estimates. Meta-analyses often answer these questions through the lens of a *random-effects* model (DerSimonian and Laird, 1986). The model assumes that each of the multiple estimates pertains to a distinct parameter value drawn at random from a population of potential parameter values. Then a weighted average of the estimates is interpreted to be an estimate of the mean of all potential parameter values.

Medical researchers have used random-effects models to perform numerous meta-analyses of studies evaluating treatments for many diseases. For example, Burchwald *et al.* (2008) used this approach to combine the findings of 134 studies of the outcomes of bariatric surgery. The 134 studies included 5 randomized trials, 28 nonrandomized but somehow otherwise controlled trials, and 101 other studies described as "uncontrolled case series." The studies were performed around the world. They followed patients for different periods of time and measured weight loss, a primary health outcome of interest, in multiple ways. To summarize the findings of the meta-analysis, the authors write (p. 1724): "The mean (95% confidence interval) percentage of excess weight loss was 61.2% (58.1%-64.4%) for all patients." The estimated mean of 61.2% considers the 134 studies to be a sample drawn from a population of potential studies. It is not clear what implications should be drawn by a clinician who treats a specific population of patients.

The relevance to clinical practice of a weighted average of estimates is often similarly obscure. DerSimonian and Laird consider each of the studies considered in a meta-analysis to be drawn at random "from a population of possible studies." They do not explain what is meant by a population of possible studies, nor why the published studies should be considered a random sample from this population. Even if these concepts are meaningful, they do not explain how a mean outcome across a population of possible

studies connects to what should matter to a clinician, namely the distribution of health outcomes across the relevant population of patients.

Medical researchers who have performed meta-analyses have struggled to explain how clinicians should use the findings. For example, Chen and Parmigiani (2007) report a meta-analysis of ten studies predicting risk of breast and ovarian cancer. The authors describe a weighted average of the risks reported by the studies as a "consensus estimate." In fact, there was no consensus across studies, which reported a range of estimates pertaining to heterogeneous populations.

### 3. Rationales for Incredible Certitude

I discuss here several potential rationales for incredible certitude that aim to enhance social welfare. These are psychological necessity, simplification of individual decision making, and coordination of beliefs across persons.

Researchers may also express certitude with private objectives in mind. They may believe that the scientific community and the public reward researchers who assert strong findings and doubt those who express uncertainty. They may conflate science with advocacy, tailoring their analyses to generate conclusions that they prefer. These private considerations may motivate some researchers, but they do not offer reasons why society should encourage incredible certitude.

#### 3.1. Psychological Necessity

Some colleagues assert that expression of incredible certitude is necessary because the consumers of research are psychologically unable or unwilling to cope with uncertainty. I have repeatedly heard this

assertion from economists who advise policymakers. They contend that, if they were to express uncertainty, policymakers would either misinterpret findings or not listen at all. This contention is nicely illustrated by the story that circulates about an economist's attempt to describe uncertainty about a forecast to President Lyndon B. Johnson. The economist is said to have presented the forecast as a likely range of values for the quantity under discussion. Johnson is said to have replied "Ranges are for cattle. Give me a number."

Beyond provision of anecdotes, colleagues may state broadly that "psychologists have shown" that humans can't deal with uncertainty, without providing citations. What has research in psychology and related fields shown about the ability and willingness of humans to deal with uncertainty? Several literatures discussed below relate to this question. They do not provide a basis to conclude that expression of incredible certitude is a psychological necessity.

### 3.1.1. Intolerance of Uncertainty

Clinical psychologists have studied "intolerance of uncertainty" (IU) as a phenomenon associated with the clinical disorder called "generalized anxiety disorder" (GAD). Buhr and Dugas (2009) define IU as follows (p. 216):

"Research has shown that intolerance of uncertainty is a fundamental cognitive process involved in excessive worry and GAD. Intolerance of uncertainty can be viewed as a dispositional characteristic that results from a set of negative beliefs about uncertainty and its implications (Dugas & Robichaud, 2007) and involves the tendency to react negatively on an emotional, cognitive, and behavioral level to uncertain situations and events (Dugas, Buhr, & Ladouceur, 2004). More specifically, individuals who are intolerant of uncertainty find uncertainty stressful and upsetting, believe that uncertainty is negative and should be avoided, and experience difficulties functioning in uncertainty-inducing situations (Buhr & Dugas, 2002). These individuals find many aspects of life difficult to tolerate given the inherent uncertainties of daily living. They tend to feel threatened in the face of uncertainty and engage in futile attempts to control or eliminate uncertainty."

If IU as defined here were a common occurrence, researchers might have good reason to think that expression of incredible certitude is a psychological necessity. However, it does not appear to be common. I am unaware of estimates of the prevalence of IU, but Kessler and Wittchen (2002) and Craske and Stein (2016) give estimates of the prevalence of GAD, a disorder that encompasses IU and much else. Relying on epidemiological surveys from various countries, they respectively report that 4–7% or 3–5% of persons suffer from GAD at some point in their lives. These estimates, to the extent that they are accurate, give upper bounds on the lifetime prevalence of IU. If the lifetime prevalence of IU is no more than 4–7% or 3–5%, the disorder is too rare for researchers to conclude that incredible certitude is a psychological necessity.

Moreover, it appears that IU is a treatable disorder when it does occur. Clinical psychologists have developed "intolerance of uncertainty therapy" (IUT) as a treatment. IUT is defined by Van der Heiden, Muris, and van der Molen (2012) as follows (p. 103): "IUT focuses on decreasing anxiety and the tendency to worry by helping patients develop the ability to tolerate, cope with, and even accept uncertainty in their everyday lives." Reporting on a randomized trial comparing IUT with other treatments for GAD, these authors report that IUT yields clinically significant reduction in patient experience of the symptoms of GAD.

### 3.1.2. Motivated Reasoning Regarding Uncertainty

Now consider the general population, the 93% or more of persons who do not have diagnosable IU disorder. Economists studying the general population have commonly maintained a sharp conceptual distinction between preferences and beliefs. This distinction is expressed particularly cleanly in the expected utility model. A utility function evaluates the desirability of an action in a specified state of nature and a subjective probability distribution expresses beliefs about the likelihood of each feasible state.

In contrast, social psychologists commingle preferences and beliefs in various ways. They sometimes use the broad term *motivated reasoning*; see the review article of Kunda (1990). Some closing of the gap between economic and social psychological thinking is evident in a small recent economic literature that

formalizes the notion of motivating reasoning. See Akerlof and Dickens (1982), Caplin and Leahy (2001), Brunnermeier and Parker (2005), Gollier and Muermann (2010), and Bénabou and Tirole (2016).

A subset of the work by social psychologists focuses on uncertainty as a motivating force per se. Bar-Anan, Wilson and Gilbert (2009) put it this way (p. 123): "Uncertainty has both an informational component (a deficit in knowledge) and a subjective component (a feeling of not knowing)." The idea of "a feeling of not knowing" has no interpretation in the expected utility model.

While social psychologists embrace the broad notion that uncertainty engenders feelings (aka affect or emotion), the literature has not attained a consensus conclusion about the nature of the feelings. Citing earlier research, Bar-Anan, Wilson, and Gilbert (2009) initially write that (p. 123) "uncertainty is generally viewed as an aversive state that organisms are motivated to reduce." This view, if accurate, might give researchers an incentive to express certitude to mitigate the negative feelings that persons obtain from uncertainty. However, these authors go on to question the general view, stating (p. 123):

"In contrast, we propose an *uncertainty intensification hypothesis*, whereby uncertainty makes unpleasant events more unpleasant (as prevailing theories suggest) but also makes pleasant events more pleasant (contrary to what prevailing theories suggest)."

The theme that uncertainty may sometimes be pleasurable is developed further in other papers including Wilson et al. (2005) and Whitchurch, Wilson, and Gilbert (2011). The latter has the unusual title "'He Loves Me, He Loves Me Not . . .': Uncertainty Can Increase Romantic Attraction."

### 3.1.3. Expression of Uncertainty in Probability Judgments

Possible evidence for the psychological view that persons are motivated to reduce uncertainty exists within a body of empirical research that asks subjects to place subjective probabilities on the truth of objectively verifiable statements and subjective distributions on the values of objectively measurable quantities. Many studies have reported findings of overconfidence. Combining evidence across multiple experiments, psychologists have found that reported subjective probabilities that statements are true tend to

be higher than the frequency with which they are true. Confidence intervals for real-valued quantities tend to be too narrow. The phenomenon has come to be called "overconfidence bias." Tversky and Kahneman (1974) and Fischhoff and MacGregor (1982) view overconfidence bias as a well-established and widespread phenomenon.

Nevertheless, the literature on overconfidence bias does not provide a rationale for scientists to express incredible certitude. The subjects in psychological experiments typically do not manifest bias so extreme as to give responses of 0 or 1 when asked to state subjective probabilities of uncertain events. They commonly give responses that express uncertainty, but perhaps not as much uncertainty as warranted. Moreover, Gigerenzer, Hoffrage, and Kleinbölting (1991) and others argue that research findings on overconfidence bias are fragile. They report that subjects often express more uncertainty when they are asked questions with different wording than psychologists have traditionally used.

Further reason to question the prevalence of overconfidence appears in the large body of economic research that elicits subjective probabilities of future personal events from survey respondents. This literature finds substantial heterogeneity in the expectations that persons hold, including the degree to which they express uncertainty. It does not find that respondents are generally overconfident. Review articles by Manski (2004, 2018c) describe the emergence of this field and summarize a range of applications. Review articles by Hurd (2009), Armantier *et al.* (2013), Delavande (2014), and Schotter and Trevino (2014) respectively focus on work measuring probabilistic expectations of older persons, inflation, populations in developing countries, and subjects making decisions in lab experiments.

### 3.2. Simplification of Individual Decision Making

A possible rationale for incredible certitude is that it may be useful as a device to simplify individual decision making under uncertainty. The broad idea, following Simon (1955), is that humans are boundedly



rational, in the sense of having computational limitations in cognition. Simon argued that it may be burdensome or infeasible for persons to make choices with the decision criteria studied in standard decision theory. Hence, he suggested that people use approximations or heuristics to reduce decision effort. I have not heard proponents of incredible certitude give this rationale explicitly, but it may perhaps underlie some thinking on the subject.

As background, I first review standard decision theory below. I then consider simplification of decision making by "as-if optimization" with incredible certitude. In general, I find it difficult to motivate as-if optimization. I offer a limited motivation that may have merit in certain decision problems.

### 3.2.1. Standard Decision Theory

The standard formalization of decision under uncertainty supposes that a decision maker must choose among a set of feasible actions. The decision maker faces uncertainty if the welfare achieved by an action varies with the *state of nature*; that is, features of the environment that are incompletely known. To begin, the decision maker lists all the states of nature that he believes could possibly occur. This list, called the *state space*, expresses partial knowledge. The larger the state space, the less the decision maker knows about the consequences of each action. Formally, Let  $C$  be the choice set and  $S$  be the state space. A welfare function  $w(\cdot, \cdot): C \times S \rightarrow \mathbb{R}^1$  maps actions and states into welfare.

The fundamental difficulty of decision making under uncertainty is clear even in a simple setting with two feasible actions and two states of nature. Suppose that one action yields higher welfare in one state of nature and the other action yields higher welfare in the other state. Then the decision maker does not know which action is better. Thus, optimization is impossible.

Decision theory suggests a two-step process, the first being obvious and the second being subtle. The first step is to eliminate *dominated* actions; that is, those which are definitely inferior to others. Formally,

action  $c \in C$  is weakly dominated if there exists a  $d \in C$  such that  $w(d, s) \geq w(c, s)$  for all  $s \in S$  and  $w(d, s) > w(c, s)$  for some  $s \in S$ .

Let  $D \subset C$  denote the subset of undominated actions. The second step is to choose an action in  $D$ . This is subtle because there is no optimal choice. There are at most various "reasonable" ways, each with its own properties.

What are reasonable ways to choose among undominated actions? When addressing this question, decision theorists have distinguished three primary situations regarding information that a decision maker may or may not have beyond specification of the state space. They have studied decision criteria suited to each situation.

In the situation with the strongest information, the decision maker asserts knowledge of an objective probability distribution on the state space, say  $P$ , generating the actual state of nature. Economists often call this knowledge *rational expectations*. The usual prescription for decision making is to maximize expected utility. The criterion is

$$(1) \quad \max_{c \in D} \int w(c, s) dP.$$

In the intermediate situation, the decision maker does not assert knowledge of an objective distribution generating the state. Instead he places a subjective distribution, say  $\pi$ , on the state space. The usual prescription is to maximize subjective expected utility. That is,

$$(2) \quad \max_{c \in D} \int w(c, s) d\pi.$$

In the situation with the weakest information, the decision maker asserts no knowledge beyond that the true state of nature lies in the specified state space. Decision theorists refer to this as a situation of *ambiguity* or *deep uncertainty*. When making a choice under ambiguity, a reasonable way to act is to use a decision criterion that achieves adequate performance in all states of nature. There are multiple ways to formalize this idea. Two commonly studied are the maximin and minimax-regret (MR) criteria.

The maximin criterion chooses an action that maximizes the minimum welfare that might possibly occur across all states of nature. The criterion is

$$(3) \quad \max_{c \in D} \min_{s \in S} w(c, s).$$

The minimax-regret criterion considers each state and computes the loss in welfare that would occur if one were to choose a specified action rather than the one that is best in this state; that is,  $\max_{d \in D} w(d, s) - w(c, s)$ . This quantity, called *regret*, measures the nearness to optimality of the action in the state. The decision maker must choose without knowing the true state. To achieve adequate performance in all states of nature, he computes the maximum regret of each action; that is, the maximum distance from optimality that the action would yield across all states. The criterion chooses an action that minimizes this maximum distance from optimality. Thus, a minimax-regret choice solves the problem

$$(4) \quad \min_{c \in D} \max_{s \in S} [\max_{d \in D} w(d, s) - w(c, s)].$$

### 3.2.2. As-If Optimization with Incredible Certitude

Standard decision theory presumes that decision makers are able to behave as prescribed. That is, they can effortlessly determine the set  $D$  of undominated actions and determine a choice that satisfies a decision

criterion such as (1)–(4). However, researchers who apply decision theory are well aware that these tasks may require substantial computational effort or be intractable.

Difficulties in determining the undominated actions are often so severe that applied decision analysts bypass the first step of the choice process. That is, they apply criteria (1)–(4) to the full choice set  $C$  rather than to the undominated subset  $D$ . When one of these criteria yields a unique choice, it necessarily is undominated. However, when a criterion yields a set of equally good choices, the set may include options that are dominated in particular ways.

The feasibility of applying criteria (1)–(4) depends on the setting, but they often become less tractable as the sizes of the choice set  $C$  and the state space  $S$  grow. Maximization of expected utility requires integration of welfare over  $S$  and then maximization over  $C$ . The maximin and minimax-regret criteria require solution of saddle point problems in  $S$  and  $C$ . The literature in applied decision analysis encounters many cases in which it is infeasible to find exact solutions to these problems, even with modern computers and software. Researchers use numerical or analytical approximations to simplify.

Expressing incredible certitude enables a more extreme simplification than is typically performed in applied decision analysis. One selects a single state of nature, say  $s^*$ , and optimizes "as if" this is the actual state. Thus, one solves the problem

$$(5) \quad \max_{c \in C} (c, s^*).$$

Computationally, this is much simpler than application of criteria (1)–(4).

The question, of course, is the quality of the decision yielded by as-if optimization. When criterion (5) yields a unique solution, the choice is necessarily undominated. However, it does not seem possible to say anything further without placing more structure on the decision problem. Depending on the circumstances, as-if optimization may yield relatively high or low expected welfare, minimum welfare, or maximum regret.

As-if optimization cannot yield some choices that may be attractive from the perspective of criteria (1)–(4). Perhaps most obviously, it cannot yield a choice that involves costly information acquisition. If one acts as if the actual state is  $s^*$ , there exists no relevant information to acquire.

As-if optimization also cannot yield diversification. Consider the classical financial problem of portfolio allocation, where an investor must allocate an endowment between two investments such as stocks and bonds. It is well-known that the allocation maximizing expected utility is a diversified portfolio when welfare is a sufficiently concave function of the investment return and when the distribution of returns has sufficient spread. Manski (2009) studies the equivalent problem of allocation of two treatments to a population and shows that the minimax-regret criterion always yields a diversified allocation under uncertainty. However, as-if optimization does not diversify. It allocates the entire endowment (population) to the investment (treatment) that gives the higher return (welfare) in state  $s^*$ .

### 3.2.3. A Limited Rationale for As-If Optimization

There exist some decision problems whose structure give a limited rationale for as-if optimization as a device for simplification of decision making. I describe one here.

Let welfare function  $w(\cdot, \cdot)$  be uniformly bounded. Without loss of generality, let the lower and upper bounds be 0 and 1. Let  $c^*$  denote an action that solves the as-if optimization problem (5). Let  $c'$  denote an action ranked immediately below  $c^*$  in terms of (5).

Suppose that one uses expected utility to evaluate decisions and places positive probability, say  $\alpha$ , on the state  $s^*$  assumed in as-if optimization. Suppose that one only computes welfare in state  $s^*$ . Then one can determine that the expected utility yielded by action  $c^*$  lies in the range  $[\alpha \cdot w(c^*, s^*), \alpha \cdot w(c^*, s^*) + (1 - \alpha)]$ . The lower bound would occur if action  $c^*$  were to yield welfare 0 in all states other than  $s^*$  and the upper bound would occur if  $c^*$  were to give welfare 1 in all other states. Similarly, the expected utility yielded by

action  $c'$  lies in the range  $[\alpha \cdot w(c', s^*), \alpha \cdot w(c', s^*) + (1 - \alpha)]$ . Indeed,  $\alpha \cdot w(c', s^*) + (1 - \alpha)$  is an upper bound on the expected utility yielded by any action other than  $c^*$ .

These findings imply an upper bound on the value of making the effort to maximize expected utility. If one were to maximize expected utility, the greatest gain that one could potentially make relative to choice of  $c^*$  is  $\max\{0, [\alpha \cdot w(c', s^*) + (1 - \alpha)] - \alpha \cdot w(c^*, s^*)\}$ . This holds because  $\alpha \cdot w(c', s^*) + (1 - \alpha)$  is the highest possible expected utility with choice of any action other than  $c^*$ , while  $\alpha \cdot w(c^*, s^*)$  is the lowest with choice of  $c^*$ . Hence, maximization of expected utility cannot improve on choice of  $c^*$  if  $\alpha \cdot w(c^*, s^*) \geq \alpha \cdot w(c', s^*) + (1 - \alpha)$ . It can improve on it by at most  $[\alpha \cdot w(c', s^*) + (1 - \alpha)] - \alpha \cdot w(c^*, s^*)$  otherwise.

Now consider the marginal cost of making the effort to maximize expected utility, measured in the same units as welfare  $w$ . Suppose that this effort cost exceeds  $[\alpha \cdot w(c', s^*) + (1 - \alpha)] - \alpha \cdot w(c^*, s^*)$ . Then maximization of expected utility cannot be worth the effort.

This rationale for as-if optimization is limited in two respects. First, the argument that as-if optimization is superior to maximization of expected utility holds only if the probability  $\alpha$  placed on state  $s^*$  and the marginal effort cost of maximizing expected utility are both sufficiently high. Second, as-if optimization may not be the only alternative to maximization of expected utility. There may exist other decision strategies that are computationally less burdensome than maximization of expected utility but yield better decisions than as-if optimization.

### 3.3. Using As-If Consensus to Coordinate Beliefs for Collective Decisions

Section 3.2 considered use of as-if optimization to simplify individual decision making. A related idea is to use "as-if consensus" to simplify collective decision making. As-if consensus means that the members of a community agree to accept a conventional certitude, which asserts that some specified state of nature

holds. The motivation is that this eliminates coordination failures that may arise if persons recognize uncertainty and deal with it in different ways.

I observed earlier that as-if optimization cannot yield a choice that involves costly information acquisition. The same hold with as-if consensus. If a collectivity acts as if the actual state is  $s^*$ , there exists no relevant information to acquire. Thus, a society that acts as if it knows the truth about a subject would not fund new research on it.

It does not seem possible to say more without placing structure on the collective decision problem. I am aware of one context with a compelling argument for as-if consensus. This is in establishment of rules for financial accounting, discussed below.

### 3.3.1. Dealing with Uncertainty in Financial Accounting

In a conversation about the CBO practice of providing Congress with only a point prediction (score) of the future impact of legislation on the federal debt, a member of the CBO staff told me that a point prediction was necessary because scores play an official role in the federal accounting system. The person noted that all financial accounting systems use point estimates of revenues, costs, and the value of assets.

The accounting literature has long been aware of uncertainties in the estimates that accounting systems make; see Brief (1975) for a historical perspective. The question has been how to deal with uncertainty. The universal answer has been to propose conventions for producing point estimates and seek to have them widely accepted, the result being as-if consensus.

As-if consensus seems essential when formulating rules for transactions. Without it, parties may not agree on the amounts to be transacted. Consider, for example, the use by the federal government of decennial state-by-state Census population estimates in apportionment of the U.S. House of Representatives and allocation of federal funds across the states. It is recognized that Census population estimates may have various forms of error; see, for example, Seeskin and Spencer (2015). Nevertheless, apportionment and fund

allocation require that the Census Bureau use some convention to produce a point estimate of each state's population.

The use of point estimates in accounting may be inevitable, but such use does not imply that the producers of these estimates should act as if they are errorless. The conceptual framework for accounting promulgated in Financial Accounting Standards Board (2010) is instructive. The framework calls for accountants to provide a "faithful representation" of financial information, writing (p. 17):

"To be useful, financial information not only must represent relevant phenomena, but it also must faithfully represent the phenomena that it purports to represent. To be a perfectly faithful representation, a depiction would have three characteristics. It would be complete, neutral, and free from error. Of course, perfection is seldom, if ever, achievable."

It goes on to say (p. 18):

"Faithful representation does not mean accurate in all respects. Free from error means there are no errors or omissions in the description of the phenomenon, and the process used to produce the reported information has been selected and applied with no errors in the process. In this context, free from error does not mean perfectly accurate in all respects. For example, an estimate of an unobservable price or value cannot be determined to be accurate or inaccurate. However, a representation of that estimate can be faithful if the amount is described clearly and accurately as being an estimate, the nature and limitations of the estimating process are explained, and no errors have been made in selecting and applying an appropriate process for developing the estimate."

I find admirable the manner in which the Board defines "free from error." It does not ask that a financial estimate or prediction be "perfectly accurate in all respects," which would require incredible certitude. It asks the accountant to describe without error "the process used to produce the reported information" and to explain "the limitations of the estimating process." Thus, the Board calls on accountants to describe uncertainty transparently rather than hide it.

I urge the CBO to do likewise when reporting scores to Congress. The CBO can continue to provide point predictions for use in the federal accounts. However, the CBO should also document clearly the



process used to produce its scores. Moreover, the CBO can accompany scores with quantitative measures of their uncertainty, which Congress may find useful as it considers pending legislation.

#### 4. Conclusion

This paper has documented incredible certitude in scientific reporting of findings, reiterating and adding to the documentation in my earlier work. Section 2 provided multiple illustrative cases of conventional certitude, dueling certitudes, and sacrificing relevance for certitude. The central new aspect of the paper is its exploration in Section 3 of several rationales that might perhaps justify expression of incredible certitude: psychological necessity, simplification of individual decision making, and coordination of beliefs across persons.

I do not find much basis in psychological and related research to conclude that humans can't cope with uncertainty and hence require incredible certitude. Rather than express incredible certitude, it would be more constructive for researchers to convey uncertainty clearly. The Bank of England provides a nice case study of graphical communication of uncertainty in its fan charts (Aikman *et al.* 2011). A body of work on communication of scientific uncertainty offers a set of suggestions that economists may find helpful. See Morgan and Henrion (199), Fischhoff (2012), and Fischhoff and Davis (2014).

I also do not find much support for the idea that as-if optimization is an effective way to simplify individual decision making. As-if optimization requires less effort relative to standard criteria such as (1)–(4) for decision making under uncertainty, but it may seriously degrade the quality of decisions. I pointed out that as-if optimization cannot yield classes of strategies that may be attractive under uncertainty, including ones that involve information acquisition or diversification. I called attention to a limited set of circumstances in which as-if optimization may be appealing, these being when a decision maker wants to

maximize expected utility, computation of expected utilities requires much effort, and high probability is placed on the state of nature used in as-if optimization.

The use of as-if consensus to coordinate beliefs across persons seems inevitable in the financial accounting systems used to make transactions in modern economies. I find it hard to envision a workable accounting system that does not use conventions to make point estimates of revenues, costs, and asset values. However, accounting aside, I find as-if consensus difficult to justify.

In toto, I find that scientific expression of incredible certitude at most has practical appeal in certain limited contexts. On principle, I view characterization of uncertainty as a fundamental aspect of the scientific code of conduct. Hence, I conclude that researchers should generally strive to convey uncertainty clearly.

## References

- Akerlof, G., and W. Dickens (1982), "The Economic Consequences of Cognitive Dissonance," *American Economic Review*, 72, 307-19.
- Aikman, D., P. Barrett, S. Kapadia, M. King, J. Proudman, T. Taylor, I. de Weymarn, and T. Yates (2011), "Uncertainty in Macroeconomic Policy-making: Art or Science," *Philosophical Transactions of the Royal Society* 369, 4798-4817.
- Aneja, A., J. Donohue, and A. Zhang (2011), "The Impact of Right-to-Carry Laws and the NRC Report: Lessons for the Empirical Evaluation of Law and Policy," *American Law and Economic Review*, 13, 565-632.
- Angrist, J., G. Imbens, and D. Rubin (1996), "Identification of Causal Effects Using Instrumental Variables," *Journal of the American Statistical Association*, 91, 444-455.
- Armantier, O., W. Bruine de Bruin, S. Potter, G. Topa, W. van der Klaauw, and B. Zafar (2013), "Measuring Inflation Expectations," *Annual Review of Economics*, 5, 273-301.
- Auerbach, A. (1996), "Dynamic Revenue Estimation," *Journal of Economic Perspectives*, 10, 141-157.
- Bar-Anan, Y., T. Wilson, and D. Gilbert (2009), "The Feeling of Uncertainty Intensifies Affective Reactions," *Emotion*, 9, 123-127.
- Barsky, R. (1998), *Noam Chomsky: A Life of Dissent*, Cambridge: MIT Press.
- Bénabou, R. and J. Tirole (2016), "Mindful Economics: The Production, Consumption, and Value of Beliefs," *Journal of Economic Perspectives* 30, 141–64.
- Berkson, J. (1958), "Smoking and Lung Cancer: Some Observations on Two Recent Reports," *Journal of the American Statistical Association*, 53, 28–38.
- Black, D. and D. Nagin (1998), "Do Right-to-Carry Laws Deter Violent Crime?" *Journal of Legal Studies*, 27, 209-219.
- Brief, R. (1975), "The Accountant's Responsibility in Historical Perspective," *The Accounting Review*, 50, 285-297.
- Brunnermeier, M. and J. Parker (2005), "Optimal Expectations," *American Economic Review*, 95, 1092–118.
- Buhr, K. and M. Dugas (2002), "The Intolerance of Uncertainty Scale: Psychometric Properties of the English Version," *Behaviour Research and Therapy*, 40, 931-945.
- Buhr, K. and M. Dugas (2009), "The role of fear of anxiety and intolerance of uncertainty in worry: An experimental manipulation," *Behaviour Research and Therapy*, 47, 215-223.

Burchwald, H., Y. Avidor, E. Braunwald, M. Jensen, W. Pories, K. Fahrbach, and K. Schoelles (2008), "Bariatric Surgery: A Systematic Review and Meta-analysis," *Journal of the American Medical Association*, 292, 1724-1737.

Bureau of Labor Statistics (2018) Employment Situation Technical Note. <https://www.bls.gov/news.release/empsit.tn.htm>, accessed July 9, 2018.

Campbell, D. (1984), "Can We Be Scientific in Applied Social Science?" *Evaluation Studies Review Annual*, 9, 26–48.

Campbell, D. and J. Stanley (1963), *Experimental and Quasi-Experimental Designs for Research*, Chicago: Rand McNally.

Caplin, A. and J. Leahy (2001), "Psychological Expected Utility Theory and Anticipatory Feelings," *Quarterly Journal of Economics*, 116, 55–80.

Chen, S. and G. Parmigiani (2007), "Meta-Analysis of *BRCA1* and *BRCA2* Penetrance," *Journal of Clinical Oncology*, 25, 1329-1333.

Congressional Budget Office (2017) Cost Estimate. American Health Care Act, <https://www.cbo.gov/system/files/115th-congress-2017-2018/costestimate/americanhealthcareact.pdf>, accessed April 5, 2017.

Cornfield, J. (1951), "A Method of Estimating Comparative Rates from Clinical Data. Applications to Cancer of the Lung, Breast, and Cervix," *Journal of the National Cancer Institute*, 11, 1269–1275.

Crane, B, A. Rivolo, and G. Comfort (1997), *An Empirical Examination of Counterdrug Interdiction Program Effectiveness*, IDA paper P-3219, Alexandria, VA: Institute for Defense Analyses.

Craske, M. and M. Stein (2016), "Anxiety," *Lancet*, 388, 3048-59.

Croushore, D. (2011), "Frontiers of Real-Time Data Analysis," *Journal of Economic Literature*, 49, 72-100.

Deaton, A. (2009), "Instruments of Development: Randomization in the Tropics, and the Search for the Elusive Keys to Economic Development," National Bureau of Economic Research Working Paper 14690.

Delavande, A. (2014), "Probabilistic Expectations in Developing Countries," *Annual Review of Economics*, 6, 1-20.

DerSimonian, R. and N. Laird (1986), "Meta-Analysis in Clinical Trials," *Controlled Clinical Trials*, 7, 177-188.

Dugas, M., K. Buhr, and R. Ladouceur (2004), "The Role of Intolerance of Uncertainty in the Etiology and Maintenance of Generalized Anxiety Disorder," in R. Heimberg, C. Turk, and D. Mennin (eds.), *Generalized Anxiety Disorder: Advances in Research and Practice*, pp. 143–163, New York: Guilford Press.

Dugas, M. and M. Robichaud (2007), *Cognitive-behavioural Treatment for Generalized Anxiety Disorder: from Science to Practice*, New York: Routledge.

- Duggan, M. (2001), "More Guns, More Crime," *Journal of Political Economy*, 109, 1086-1114.
- Durlauf, S, S. Navarro, and D. Rivers (2016), "Model Uncertainty and the Effect of Shall-Issue Right-to-Carry Laws on Crime," *European Economic Review*, 81, 32-67.
- Elmendorf, D. (2010), letter to Honorable Nancy Pelosi, Speaker, U. S. House of Representatives, Congressional Budget Office, [http://housedocs.house.gov/energycommerce/hr4872\\_CBO.pdf](http://housedocs.house.gov/energycommerce/hr4872_CBO.pdf), accessed April 5, 2017.
- Duignan, B. (2017), "Occam's Razor," *Encyclopædia Britannica*, <https://www.britannica.com/topic/Occams-razor>. Accessed July 9, 2018.
- Financial Accounting Standards Board (2010), *Statement of Financial Accounting Concepts No. 8*, [www.fasb.org/jsp/FASB/Document\\_C/DocumentPage?cid=1176157498129&acceptedDisclaimer=true](http://www.fasb.org/jsp/FASB/Document_C/DocumentPage?cid=1176157498129&acceptedDisclaimer=true), accessed August 3, 2018.
- Fischhoff, B. (2012), "Communicating Uncertainty: Fulfilling the Duty to Inform," *Issues in Science and Technology*, 28, 63-70.
- Fischhoff, B. and A. Davis (2014), "Communicating Scientific Uncertainty," *Proceedings of the National Academy of Sciences*, 111, 13664-13671.
- Fischhoff, B. and D. MacGregor (1982), "Subjective Confidence in Forecasts," *Journal of Forecasting*, 1, 155-172.
- Fixler, D. R. Greenaway-McGrevy, and B. Grimm (2011), "Revisions to GDP, GDI, and Their Major Components," *Survey of Current Business*, 91, 9-31.
- Fixler, D., R. Greenaway-McGrevy, and B. Grimm (2014), "Revisions to GDP, GDI, and Their Major Components," *Survey of Current Business*, 94, 1-23.
- Fleiss, J. 1981. *Statistical Methods for Rates and Proportions*. New York: Wiley.
- Friedman, M. (1953), *Essays in Positive Economics*, Chicago: University of Chicago Press.
- Gail, M, L. Brinton, D. Byar, D. Corle, S. Green, C. Shairer, and J. Mulvihill (1989), "Projecting Individualized Probabilities of Developing Breast Cancer for White Females Who Are Being Examined Annually," *Journal of the National Cancer Institute*, 81, 1879-86.
- Gigerenzer, G., U. Hoffrage, and H. Kleinbölting (1991), "Probabilistic Mental Models: A Brunswikian Theory of Confidence," *Psychological Review*, 98, 506-528.
- Heckman, J. and S. Urzua (2009), "Comparing IV with Structural Models: What Simple IV Can and Cannot Identify," National Bureau of Economic Research Working Paper 14706.
- Horowitz, J. and C. Manski (1998), "Censoring of Outcomes and Regressors due to Survey Nonresponse: Identification and Estimation Using Weights and Imputations," *Journal of Econometrics*, 84, 37-58.

- Hurd, M. (2009), "Subjective Probabilities in Household Surveys," *Annual Review of Economics*, 1, 543-564.
- Hsieh, D., C. Manski, and D. McFadden (1985), "Estimation of Response Probabilities from Augmented Retrospective Observations," *Journal of the American Statistical Association*, 80, 651-662.
- Imbens, G. (2010), "Better LATE Than Nothing: Some Comments on Deaton (2009) and Heckman and Urzua (2009)," *Journal of Economic Literature*, 48, 399-423.
- Imbens, G. and J. Angrist (1994), "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467-476.
- Imbens, G. and C. Manski (2004), "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72, 1845-1857.
- Kessler, R. and H. Wittchen (2002), "Patterns and Correlates of Generalized Anxiety Disorder in Community Samples," *Journal of Clinical Psychiatry*, 63 (suppl 8), 4-10.
- Knutti, R., R. Furrer, C. Tebaldi, J. Cermak, and G. Meehl (2010), "Challenges in Combining Projections from Multiple Climate Models," *Journal of Climate*, 23, 15 May, 2739-2758.
- Kunda, Z. (1990), "The Case for Motivated Reasoning," *Psychological Bulletin*, 108, 480-498.
- Kuznets, S. (1948), "Discussion of the New Department of Commerce Income Series," *Review of Economics and Statistics* 30, 151-179.
- Lott, J. (2010), *More Guns, Less Crime: Understanding Crime and Gun-Control Laws*, Chicago: University of Chicago Press.
- Lott, J. and D. Mustard (1997), "Crime, Deterrence and Right-to-Carry Concealed Handguns," *Journal of Legal Studies*, 26, 1-68
- Manski, C. (1989), "Anatomy of the Selection Problem," *Journal of Human Resources*, 24, 343-360.
- Manski, C. (1994), "The Selection Problem," *Advances in Econometrics, Sixth World Congress*, ed C. Sims, Cambridge: Cambridge University Press, 143-170.
- Manski, C. (1995), *Identification Problems in the Social Sciences*, Cambridge: Harvard University Press.
- Manski, C. (1996), "Learning about Treatment Effects from Experiments with Random Assignment of Treatments," *Journal of Human Resources*, 31, 707-733.
- Manski, C. (2003), *Partial Identification of Probability Distributions*, New York: Springer-Verlag.
- Manski, C. (2004), "Measuring Expectations," *Econometrica*, 72, 1329-76.
- Manski, C. (2007), *Identification for Prediction and Decision*, Cambridge: Harvard University Press.

- Manski, C. (2009), "Diversified Treatment under Ambiguity," *International Economic Review*, 50, 1013-1041.
- Manski, C. (2011), "Policy Analysis with Incredible Certitude," *The Economic Journal*, 121, F261-F289.
- Manski, C. (2013), *Public Policy in an Uncertain World*, Cambridge: Harvard University Press.
- Manski, C. (2015), "Communicating Uncertainty in Official Economic Statistics: An Appraisal Fifty Years after Morgenstern," *Journal of Economic Literature*, 53, 631-653.
- Manski, C. (2016), "Credible Interval Estimates for Official Statistics with Survey Nonresponse," *Journal of Econometrics*, 191, 293-301.
- Manski, C. (2018a), "Communicating Uncertainty in Policy Analysis," *Proceedings of the National Academy of Sciences*, forthcoming.
- Manski, C. (2018b), "Reasonable Patient Care under Uncertainty," *Health Economics*, 1-25, <https://doi.org/10.1002/hec.3803>.
- Manski, C. (2018c), "Survey Measurement of Probabilistic Macroeconomic Expectations: Progress and Promise," *NBER Macroeconomics Annual*, 32, 411-471.
- Manski, C. and J. Pepper (2018), "How Do Right-to-Carry Laws Affect Crime Rates? Coping with Ambiguity Using Bounded-Variation Assumptions," *Review of Economics and Statistics*, 100, 232-244.
- Marcus, R. (2016) Welcome to the Post-Truth Presidency. *Washington Post*, December 2, [https://www.washingtonpost.com/opinions/welcome-to-the-post-truth-presidency/2016/12/02/baaf630a-b8cd-11e6-b994-f45a208f7a73\\_story.html?utm\\_term=.650f5713beb9](https://www.washingtonpost.com/opinions/welcome-to-the-post-truth-presidency/2016/12/02/baaf630a-b8cd-11e6-b994-f45a208f7a73_story.html?utm_term=.650f5713beb9), accessed December 8, 2016.
- Mastrandrea, M., C. Field, T. Stocker, O. Edenhofer, K. Ebi, D. Frame, H. Held, E. Kriegler, K. Mach, P. Matschoss, G. Plattner, G. Yohe, and F. Zwiers (2010), *Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties*, Intergovernmental Panel on Climate Change (IPCC).
- McFarlane, N. (2011), "Parameterizations: representing key processes in climate models without resolving them," *WIREs Climate Change*, 2, July/August, 482-497.
- McGuffie, K. and A. Henderson-Sellers (2005), *A Climate Modelling Primer – Third Edition*, Chichester, England: John Wiley & Sons.
- McWilliams, J. C. (2007), "Irreducible imprecision in atmospheric and oceanic simulations," *Proceedings of the National Academy of Sciences*, 104, 8709-8713.
- Morgan, M. and M. Henrion (1990), *Uncertainty: A Guide to Dealing With Uncertainty in Quantitative Risk and Policy Analysis*, Cambridge: Cambridge University Press.
- Morgenstern, O. (1963), *On the Accuracy of Economic Observations: Second Edition*, Princeton: Princeton University Press.

- National Cancer Institute (2011), *Breast Cancer Risk Assessment Tool*, <http://www.cancer.gov/bcrisktool/>, accessed July 9, 2018.
- National Comprehensive Cancer Network (2017), *Breast Cancer Screening and Diagnosis*, Version 1.2017, [www.nccn.org/professionals/physician\\_gls/pdf/breast-screening.pdf](http://www.nccn.org/professionals/physician_gls/pdf/breast-screening.pdf), accessed March 8, 2018.
- National Research Council (1999), *Assessment of Two Cost-Effectiveness Studies on Cocaine Control Policy*, Committee on Data and Research for Policy on Illegal Drugs, Charles F. Manski, John V. Pepper, and Yonette Thomas, editors. Committee on Law and Justice and Committee on National Statistics, Commission on Behavioral and Social Sciences and Education, Washington, DC: National Academy Press.
- National Research Council (2005), *Firearms and Violence: A Critical Review*, edited by C. Wellford, J. Pepper, and C. Petrie, Washington, DC: National Academy Press.
- Palmer, T., G. Shutts, R. Hagedorn, F. Doblas-Reyes, T. Jung, and M. Leutbecher (2005), "Representing Model Uncertainty in Weather and Climate Prediction," *Annual Review of Earth and Planetary Sciences*, 33, 163-193.
- Parker, W. (2006), "Understanding Pluralism in Climate Modeling," *Foundations of Science*, 11, 349-368.
- Parker, W. (2013), "Ensemble modeling, uncertainty and robust predictions," *WIREs Climate Change*, 4, 213-223.
- Rosenbaum, P. (1999), "Choice as an Alternative to Control in Observational Studies," *Statistical Science*, 14, 259-304.
- Rydell, C. and S. Everingham (1994), *Controlling Cocaine*, Report prepared for the Office of National Drug Control Policy and the U. S. Army, Santa Monica, CA: RAND Corporation.
- Schotter, A., and I. Trevino (2014), "Belief Elicitation in the Laboratory," *Annual Review of Economics*, 6, 103-128.
- Seeskin, Z. and B. Spencer (2015), "Effects of Census Accuracy on Apportionment of Congress and Allocations of Federal Funds," Institute for Policy Research Working Paper 15-05, Northwestern University.
- Simon, H. (1955), "A Behavioral Model of Rational Choice," *Quarterly Journal of Economics*, 69, 99-118.
- Swinburne, R. (1997), *Simplicity as Evidence for Truth*, Milwaukee: Marquette University Press.
- Stainforth, D., M. Allen, E. Tredger, and L. Smith (2007), "Confidence, uncertainty and decision-support in climate predictions," *Philosophical Transactions of the Royal Society*, 365, 2145-2161.
- Tukey, J. (1962), "The Future of Data Analysis," *Annals of Mathematical Statistics*, 33, 1-67.
- Tversky, A. and D. Kahneman (1974), "Judgement Under Uncertainty: Heuristics and Biases," *Science*, 185, 1124-1131.



U.S. Census Bureau (2011), *Current Housing Reports, Series H150/09, American Housing Survey for the United States: 2009*, Washington, DC: U.S. Government Printing Office.

Van der Heiden, C., P. Muris, and H. van der Molen (2012), "Randomized Controlled Trial on the Effectiveness of Metacognitive Therapy and Intolerance-of-uncertainty Therapy for Generalized Anxiety Disorder," *Behaviour Research and Therapy*, " 50, 100-109.

Whitchurch, E., T. Wilson, and D. Gilbert (2011), "'He Loves Me, He Loves Me Not . . .': Uncertainty Can Increase Romantic Attraction," *Psychological Science*, 22, 172-175.

Wilson, T., D. Centerbar, D. Kermer, and D. Gilbert (2005), "The Pleasures of Uncertainty: Prolonging Positive Moods in Ways People Do Not Anticipate," *Journal of Personality and Social Psychology*, 88, 5-21.