Confirmation bias in government-sponsored policy evaluation: A conceptual framework

Arnaud Vaganay

Meta-Lab Working Paper 1
October 2016

This paper investigates a little-understood question in political science: to what extent do political institutions influence policy evaluation?

A large literature has flourished in recent years to analyse how institutions affect each stage of the policy cycle, which commonly includes agenda setting, policy formulation, implementation, and evaluation. Thanks to these contributions, we have learned much about the reasons and the ways in which institutions produce sub-optimal policies or 'policy bias' by favouring one group or issue to the detriment of others [1], [2]. For instance, agenda-setting theories have shown how these institutions compete for turning private issues into public policy. Other theories have emphasised the key role played by these different institutions in the formulation of policy. We have also learned much about how institutions can affect the implementation of reforms.

Comparatively, our knowledge of what influences the way policy evaluations are conducted appears limited. The authors of *The Oxford Handbook of Political Institutions* barely address the issue. Other textbooks are equally succinct [4]–[6]. Yet, policy evaluations raise important questions, mainly because they tend to be conducted or commissioned by the very organisations that designed and implemented the intervention in the first place. The assumption of independence, which underpins scientific research, is often violated.

Political institutions typically include formal democratic bodies (parliaments, governments, bureaucracies, political parties, presidents, etc.), however, institutional theories have also scrutinised the role groups and organisations without a constitutional mandate but nevertheless influential (interest groups, media, pollsters). The institutions I will be referring to in this paper are essentially government departments and agencies.

I define policy evaluation as "the ex post assessment of the strengths and weaknesses of public programs and projects" [7]. The emphasis on ex post means that this paper does not address the literature on ex ante analysis, where methods to evaluate policy alternatives are used as decision-making aids [7]–[9]. Policy evaluation is akin to research and development (R&D) in the social sphere; with the difference that policy research is non-proprietary and can be conducted and replicated outside of government. Both policy evaluation and R&D differ from basic

research, which is experimental or theoretical work undertaken primarily to acquire new knowledge of the underlying foundation of phenomena and observable facts, without any particular application or use in view (OECD Glossary of Statistical Terms).

It should be said right away that political institutions influence research in two capacities. First, as regulators of scientific activities: throughout the 20th century, political institutions have increasingly influenced research notably through public funding, the sanctioning of research misconduct and the definition of the research agenda. Second, institutions influence research as clients of scientific organisations. This paper focuses on this second aspect only.

The goal of this paper is to lay the foundations for an empirical research agenda assessing the effect of political institutions on policy evaluation. Three specific objectives have been assigned to it. First, this paper reviews the theoretical and empirical literature on the interaction between science and institutions – the last review dating back from 1998 [10]. This review will in turn help me identify the type of issues that would make a significant contribution to this scholarship. The second objective of this paper is to come up with a 'better' theory. The third objective is to operationalise the research question.

The rest of this paper is structured as follows. The first section synthesises what we know about the effect of political institutions on research and identifies gaps in the literature. The second section offers an alternative approach to this scholarship based on the idea of 'confirmation bias' and briefly introduces this idea. The third section operationalises this approach. The fourth and fifth sections consider different independent and dependent variables respectively. The sixth section discusses the specificities of social research compared with other types of applied research such as clinical trials. The final section concludes.

State of the literature

The literature on the effect of political institutions on policy research spans several disciplines, including philosophy, research methods, political science, sociology and diverse 'professional' literatures including education and nursing. A systematic review of the literature would be a difficult exercise. Rather, the following section reviews the literature in a *narrative* fashion, focusing on what I considered to be the most significant contributions and highlighting points of consensus and disagreement.

Research without institutions

Assessing the effect of political institutions on policy research requires a thought experiment, namely the identification of the principles guiding research in a state of nature, or more realistically, in a context where scientists would work with minimum constraints. Other things remaining equal, any deviation from these principles occurring in an institutional context can be attributed to these very institutions.

The history and philosophy of science argue that these principles have been defined in two phases. Until the Enlightenment, science was primarily defined by its *purpose*, namely the advancement of knowledge. As such, it was virtually undistinguishable

from philosophy. In a state of nature, research would be conducted by free individuals pursuing neither private gain nor political ideology, but simply the truth [11]. The advent of the 'scientific revolution' – between the Renaissance and the 18th century – has led to a redefinition of science based on its *methods*. The Oxford English Dictionary defines the scientific method as "a method or procedure that has characterized natural science since the 17th century, consisting in systematic observation, measurement, and experiment, and the formulation, testing, and modification of hypotheses"[1]. The emphasis here is on the *procedure*: what makes a claim scientific is not its substance but the way the information was gathered, analysed and interpreted. Thus, theoretically, disagreements among scientists are not concerned with the relevance of the findings but with the credibility of the research process and the assumptions underpinning it [11].

In addition to these 'technical prescriptions', science is based on a number of 'moral prescriptions', which are equally binding, not because they are procedurally efficient, but because they are believed right and good [12]. These moral norms all relate to scientists' attitudes and behaviours in relation to each other and their research [13]. According to Merton, these norms include:

- *Communality* ("communism" in the original text), i.e. the common ownership of scientific results and methods and the consequent imperative to share both freely.
- *Universalism* specifies that scientific work and findings should be evaluated on the basis of "pre-established impersonal criteria: consonance with observation and with previously confirmed knowledge", and not on the personal, social or cultural attributes of the scientists involved.
- The principle of *organised scepticism* refers to the "detached scrutiny of beliefs in terms of empirical and logical criteria". This principle has implications for both producers and consumers of scientific findings: the former need to present their findings and methods transparently so that their value can be assessed, and the latter need to suspend judgement until they have examined findings and methods according to accepted standards and criteria.
- Finally, *disinterestedness* demands that scientists' work remain uncorrupted by self-interested motivations. It precludes the pursuit of science for the sake of riches, though Merton recognised the powerful influence of competition for scientific priority.

Thus, the assumption in much of the literature is that, in a state of nature, or in a state of minimum constraints, science would be independent and pursued for the sole purpose of human enlightenment. It would also scrupulously apply the scientific method and follow the moral norms of science.

Political institutions as 'consumers' of scientific advice

The fundamental difference between research undertaken in a state of nature and research undertaken in an institutional or professional context is that, in the latter situation, scientists are employed (or commissioned or compensated) and that their research is actually *utilised*.

[1] (Oxford English Dictionary, definition for 'scientific')

[2] Phase 1: Screening for safety; Phase 2: testing the efficacy of the drug, usually against a placebo; Phase 3: confirmatory study; Phase 4: post-marketing studies delineating additional information on the

The literature on 'research utilisation' has shed light on the three properties that make evidence a highly sought after resource [14]–[17]. The *conceptual* property of research is the closest to the purpose of science in a state of nature. It emphasises its capacity to enlighten individuals and organisations by articulating concepts and changing their understanding of natural and social phenomena. The *instrumental* property of research makes it capable of assisting decision-making by bringing answers to clearly predefined problems. Lastly, the *symbolic* property of research involves using research results and processes to legitimate and sustain pre-determined positions. This typology is now widely accepted in the literature and has been applied to many different policy areas including drug policy [18], urban health [19] and education [20].

Which form of utilization is most prevalent is difficult to establish given the lack of commonly agreed indicators as well as the normative aspect of the question which might bias survey responses. Though very limited, the evidence would suggest that the conceptual use of research is more prevalent in the day-to-day professional activity of professionals and managers in government agencies than symbolic utilization, which, in turn, is more important than instrumental utilization [21]. Importantly, research shows that the three types of research utilization are not mutually exclusive and are in fact frequently combined [21].

The type of research utilization (or the type of combination) depends on a number of factors. One of them is the policy area. For example, Carpenter has shown that bureaucratic agencies of state were more involved in the provision and regulation of health policy than in other policy areas [22]. The type of research utilisation depends also on the reputation that a given agency wants to enhance. Indeed, agency reputation shapes administrative decisions [23]–[25]. The literature has also convincingly shown that policy-makers are more likely to make an instrumental use of scientific advice when there is a consensus among experts on a causal theory [26]–[28].

Lastly, the literature has shed light on the aspects of research that can be used [10]. The findings are at the core of the instrumental type. In the conceptual type, utilization extends to the general ideas and generalizations from evaluation, even if they do not serve a specific and immediate purpose. In addition to the above, a policymaker using research symbolically can also take advantage of other parts of the research process. The sheer fact that an evaluation is being conducted can be used to demonstrate policy-makers' rationality and sound management [29]. Likewise, the definition of the scope of the research can be an indicator of symbolic use of evidence. Excluding inconvenient questions, areas or stakeholders from a study can help producing 'congenial' results. There is some evidence that governments can 'play it safe' when they commission an evaluation [30]. Similarly, the design of the study and the choice of measures can also be the subject of political struggles among different agencies [10], [31].

Research as a "shadow institution"

The public policy literature argues that research primarily benefits elected policy-makers (e.g. ministers). Thus, these policy-makers will seek to influence the course of research. This literature makes extensive references to 'politicians', 'political

actors' and 'interest groups'. According to this scholarship, political actors tend to frame evidence in a way that supports their agenda. This is evident is statements such as "evaluation is the continuation of politics by other means" [7].

Public policy scholars argue that democratic institutions provide incentives to successfully pass reforms. When the enactment of legislation or the implementation of a decision looks difficult, the authority of the government gets undermined, which in turn jeopardizes future reforms. The moral authority of science can, on occasions, facilitate reforms. Some have mentioned that in the US, pilot evaluations were used as "shadow institutions" used to legitimate contentious reforms [32], [33]. Thus, many have empirically sought associations between the degree of salience of an issue and the way research will be used to support policy. Some have argued that "on the small issues, evidence sometimes counts"; however, when it comes to the big issues, "politics is the order of the day" [34]. Heavily politicised policy areas are characterised by more *ad-hoc* or muddled-through policy-making [35]. In such instances, there is intense media scrutiny of decision-making and prolonged conflict between competing interest groups and a permeating sense of crisis. Typical example would be drugs, where evidence is used symbolically [36] and schools [37], [38].

Others have shown that political institutions reward politicians not so much for what they have achieved but rather for 'winning the argument'. A less demanding version of this theory suggests that policy-makers are driven less by the desire to get credit for what they have done than by the desire to avoid blame [39], [40]. The ominous label of 'failure' or 'fiasco' that hovers over policies that failed to deliver entails a political statement [7]. Thus 'cherry-picked' information can be used as 'political ammunition' in the political debate.

Regardless of the initial motivation, political institutions will occasionally lead politicians to use research symbolically, or to 'frame' it in a way that suits their aims. "They will produce – or engage others to produce – accounts of policy episodes that are, however subtly, framed and timed to convey certain ideas about what happened, why and how to judge this, and to obscure or downplay others" [7]. Methodologically, these accounts have mainly relied on individual case studies. Some of these case studies have considerably improved our understanding of the interaction between science and political institutions in areas such as drugs [36], education [37], [38]. There have been occasional reports of ministers stepping in the middle of a research project and 'leaning on' researchers [30].

Research and bureaucracy

Public administration specialists diverge from their public policy colleagues in four important ways.

First, they object the idea of a direct interest of elected policy-makers in research. According to them, elected policy-makers frame public policy, but do not carry it out. Their hypothesis is that the effect of political institutions on research is in fact mediated by government agencies and their employees. This is supported by numerous accounts of the role of expertise in the development of national agencies including the Forest Service, the Department of Agriculture [11] and the FDA [24], [25], [41]. In this context, civil servants sometimes invoke ministers to deflect blame. Some evaluators reported that opposition to evaluation would typically come from

civil servants, even though they might have pretended there was opposition from ministers [42].

Second, public administration scholars contend that research is not used with a view to secure a reform or for argumentative purposes, but to enhance the power and the legitimacy of the agency. As demonstrated by Weber, bureaucracies assert power through specialised expertise and control of information, justified by their claim to be the only means by which the complexities of modern society can be managed [11], [43]. According to Carpenter, the technical-scientific reputation of an agency is one of four reputational strategies used by public administrators to face the challenges of modern governance [23]. The result has been an ever-expanding application of administrative rationalism: seeking, with the guidance of technical expertise, rational and efficient solutions to the problems of society [11].

Third, the public administration literature is more specific when it comes to defining the effect of political institutions. At a 'macro' level, reputation gives agencies more autonomy, in the sense of being able to sway the wishes of elected officials on particular matters of policy and to secure deference from these elected officials [25], [44], [45]. Political institutions also impact the research process in very specific ways. For example, several studies have shown the timing of new drug approval by the U.S. Food and Drug Administration was influenced by the type of reputation of the agency [41], [46] as well as the relative importance of the drug, measured in terms of therapeutic novelty and expected return on investment [47]. Political institutions also influence the methodologies used in programme evaluations [31].

The fourth difference of the public administration literature is on methods. Whereas interviews, survey and desk reviews dominate the public policy literature, a more systematic approach based on administrative data has been used in public administration [25], [41], [45], [46], [48]. This is a key difference, given that only a systematic approach can reliably assess the long-term effect of political institutions on policy research.

The norms of science

It would be incomplete to review the literature on the effect of institutions on research without mentioning the response of the scientific community to the uses and misuses of science. Sociologists of science and sociologists of professions contend that the professionalization of science has led to the creation of scientific institutions. These institutions include universities, academic journals and professional societies, which all play an important role in diffusing and enforcing scientific norms.

The 'optimistic' view is that these institutions play a key role in repressing research misconduct and questionable research practices [49]. Research is a professional activity. As such it is subject to norms, i.e. prescriptions commonly known and used by practitioners [50], [51]. These prescriptions refer to which actions are required, prohibited or permitted in specific situation. The existence of such norms is a vital part being a profession. Their enforcement depends on the provision of incentives, which are reflected in the criteria used to appoint, evaluate, and promote individual faculty members. Today, the rewards of a successful academic career typically include the personal gratification derived from scholarship and discovery, recognition

by peers, and academic promotion and tenure, as well as enhanced responsibility and outside financial opportunities.

A more pessimistic view is that these institutions defend the vested interests of researchers. According to sociology of scientific knowledge, science is neither exceptional nor immune from the forces that affect other human activities. Mitroff (1974) for example, showed that for each of the Mertonian norms there exists counternorms that play equally important roles in the practice of science. Social and historic studies demonstrate science to be an enterprise consisting of individuals who passionately engage in value-laden activities to demonstrate their correctness and depend upon the socio-cultural context from which their work emanates [53]–[55]. The practice of science also includes fraudulent activity, sometimes involving even mainstream scientists (Sapp, 1990). Eventually, sociologists of science concluded that Merton's case cannot be convincingly made and that his norms might be better viewed as an ideology of science [56]. The studies of scientific practice cited above corroborate Mulkay's assertion that the Mertonian norms are an ideology that serves the interests of the scientific community in at least three ways. The norms (1) enhance the epistemic status of scientific knowledge; (2) increase the political power of scientists; and (3) elevate the social status of scientists. Functionally, they work at the interface between the scientific community and the general population and provide justification for the continued support of science in society.

Confirmation bias

Despite the above-mentioned merits, the current theoretical framework available to explain the effect of institutions on policy research is too fragmented to allow progress. There is scope for a more parsimonious and 'universal' theory.

A promising way of bridging the above-mentioned gap can be found in two related literatures. The research methods literature has approached the question as an example of 'experimenter's bias'. The experimenter's bias – also known as research bias – has been defined as "a subjective bias towards a result expected by the human experimenter" [57]. The social psychology literature has developed the related concept of 'confirmation bias' (also called confirmatory bias or 'myside' bias) which is attributed to English psychologist Peter Wason [58] and describes the tendency of people to favour information that confirms prior beliefs or hypotheses, regardless of whether the information is true [59]. The following section outlines the causes and the consequences of confirmation bias.

Causes

Confirmation bias is often described as a result of automatic, unintentional strategies rather than deliberate deception [60], [61]. It results from the supposed inability of a human being to be objective, and more specifically from (i) the desire to appear consistent and/or to fulfil public commitments; and (ii) reciprocation.

A first cause of bias is commitment. A commitment is a public engagement or obligation to take a specific course of action. Its normative power is such that individuals will often stick to the original deal even though it has changed for the

worse. The reason people stick to their commitment is that they want to maintain a positive self-image. People strengthen their original commitment by the addition of supportive new thoughts and feelings [62]. This is particularly relevant in political contexts, where reasoning can be subconsciously biased, favouring conclusions that governments have already committed to. A two-decade study of political pundits by Tetlock found that, on the whole, their predictions were not much better than chance. Tetlock divided experts into 'foxes' who maintained multiple hypotheses, and 'hedgehogs' who were more dogmatic. In general, the hedgehogs were much less accurate. Tetlock blamed their failure on confirmation bias – specifically, their inability to make use of new information that contradicted their existing theories [63]. Evidence of confirmation bias has also been found in scientific decisions [61], [64]–[66].

Another cause of bias is reciprocity. As a social construct, reciprocity means that in response to friendly actions, people are frequently much nicer and much more cooperative than predicted by the self-interest model. Conversely, in response to hostile actions they are frequently much more nasty and even brutal [67], [68]. Reciprocity is so strong that people tend to reciprocate regardless of whether they like the person who originally gave the favour and even if they did not want the favour, as was demonstrated in Regan's experiment [69]. Reciprocation can be genuine and unconscious [67]. The problem is a growing concern in the medical research community, where a vast majority of pharmaceutical companies resort to 'friendly actions'. Those include free drug samples, meals, continued medical education, financial incentives to participate in clinical trials, honoraria for delivering lectures, leisure trips, expensive text books and items of low monetary value such as pens and notepads. Reciprocity can extend to any action perceived as a 'favour' such as a job offer, a promotion, a bonus, professional honours and the sponsorship of research projects and scientific meetings [70].

There is no direct evidence of reciprocation; however there is strong evidence that scientists attitude towards industry-funded research becomes more positive as the amount of interactions between the two spheres increases [71]. A review of 17 surveys on the attitudes of researchers to financial ties in research revealed that investigators are concerned about the impact of financial ties on choice of research topic, research conduct and publication, but this concern is less among investigators already involved with industry. Researchers approve of industry collaboration and financial ties when the ties are indirectly related to the research, disclosure is up front, and results and ideas are freely publicized. However, their trust in disclosure as a way to manage conflicts may reveal a lack of awareness of the actual impact of financial incentives on themselves and other researchers [72].

Consequences

Confirmation bias leads to unconventional or sub-standard decisions. In a research context, this means decisions – and thus, potentially, findings – that are likely to be controversial within the scientific community and portrayed as not credible [11]. Opposing interests will highlight uncertainties in the evidence, discrepancies and ambiguities in the interpretation, ties between researchers and business or political interests, and any other technical aspects of the problem that can provide an opportunity to question the credibility of the research process. In other words, confirmation bias can damage scientific credibility, understood as the extent to which

science in general is recognized as a source of reliable information about the world [11]. Empirically, confirmation bias has been found to affect the way we process information, report results and interpret findings.

First, confirmation bias impairs the way we process information. Experiments have found repeatedly that people tend to test hypotheses in a one-sided way, by searching for evidence consistent with their current hypothesis [64], [73]. More specifically, confirmation bias has been invoked to explain 'illusory correlations', which is the tendency to see non-existent correlations in a set of data [74]. For example, a study recorded the symptoms experienced by arthritic patients, along with weather conditions over a 15-month period. Nearly all the patients reported that their pains were correlated with weather conditions, although the real correlation was zero [75].

Second, confirmation bias skews analyses towards an outcome that is favourable to the experimenter. The most striking case of confirmation bias is when two opposing experimenters find themselves at odds with the published findings of research they sponsor. But the Experimenter's bias is not always that spectacular. Often, it will lead to overestimate the effect of the intervention and lift any doubt regarding its possible inefficacy. Experimenter's bias has been found in very different disciplines and research areas, including in studies on the effects of nicotine [76]; antidepressants [77], [78].

Confirmation biases are not limited to the search and collection of evidence. Even when two individuals are given the same information, the way they interpret it can be biased. This has been recently demonstrated in a study on the neural responses of 30 committed partisans during the U.S. Presidential election of 2004. The authors presented subjects with reasoning tasks involving judgments about information threatening to their own candidate, the opposing candidate, or neutral control targets [79].

Mediating variables

Research shows that the strength of confirmation bias depends on the issue being considered, but not on personal differences such as cognitive abilities.

First, confirmation bias is stronger for emotionally charged issues and for deeply entrenched beliefs. This was exemplified in the *Stanford Biased Interpretation Experiment* in which participants with strong opinions about the death penalty read about mixed experimental evidence. Twenty-three percent of the participants reported that their views had become more extreme, and this self-reported shift correlated strongly with their initial attitudes [80]. More recently, Taber and Lodge conducted a similar study using the emotionally charged topics of gun control and affirmative action. They measured the attitudes of their participants towards these issues before and after reading arguments on each side of the debate. Two groups of participants showed attitude polarization: those with strong prior opinions and those who were politically knowledgeable [81].

Second, individual characteristics do not seem to have an effect on the severity of confirmation bias as previously thought. Empirical research has consistently shown that confirmation bias is persistent, regardless of intelligence level. In two

experiments involving a total of over 1400 university students and eight different comparisons, the authors found very little evidence that participants of higher cognitive ability displayed less confirmation bias [82]. There is moderate correlations between cognitive ability and the ability to avoid such biases [83]–[85].

Confirmation bias in policy research

The above shows that the question of the effect of institutions on policy research would be more effectively addressed by analysing the prevalence and severity of confirmation bias in policy research. This entails two interesting questions. The first question is normative and relates to the ideal of research. It could be formulated as follows: *To what extent is institutional policy research scientifically credible?* The second question is positivist and relates to the effect of confirmation bias: *Are evaluations of interventions to which institutions are committed less credible?*

Framing the problem in such a way would provide a number of benefits. First, confirmation bias offers a more parsimonious explanation of the effect of institutions on policy research than existing theories. It simply implies that commitments and reciprocation lead to substandard or unconventional research decisions. Second, confirmation bias works with a broad range of institutions (scientific institutions, government institutions, political institutions, private-sector companies, etc.) and policy areas. Third, it allows more accurate explanations and predictions than the idea of 'research utilisation' which many authors have struggled to operationalise.

Conveniently, the concept of confirmation bias sits comfortably with existing political science theories. First, confirmation bias leads to use research *symbolically*, i.e. for confirming the idea that the experimenter wishes to promote [15], [17], [86]–[88]. However, the concept of confirmation bias is more complete than that of research utilisation, since it offers an entire causal theory regarding the effect of political institutions on policy research. Second, confirmation bias can be seen as a type of agency cost. Governments evaluate policies on behalf of the citizenry (the principal). However, because the two parties have different interests and the government has more information (policy and research expertise), citizens cannot directly ensure that their agent is always acting their best interests [89]. Third, confirmation bias is motivated by the desire to avoid blame [40]. Were governments not sanctioned for their performance in office (in terms of curbing crime, unemployment, illiteracy, etc.), it can be argued that 'inconvenient' evaluation findings would not be a problem. Likewise, confirmation bias can be seen as the expected behaviour of organisations seeking to enhance their performative reputation [24].

Policy commitment and confirmation bias

Estimating the risk of confirmation bias can be done in different ways. The following section first describes what the sponsorship of a study can tell us about the objectivity of the researcher. It argues that the strength of the government's commitment to the intervention might be a better option.

Specifications for a correlational study

Identifying causal mechanisms, to the extent that they exist, is methodologically challenging as it requires an experimental setup. In an ideal experiment, a sample of researchers would be selected from the population and a fraction of this group – randomly selected – would be placed in political institutions. Conversely, the rest of the group would conduct the same research independently. Because the two groups are comparable by virtue of the random assignment, any systematic difference in the ways both sub-groups make policy research decisions could be attributable to political institutions. No significant difference between the incentivised group and the non-incentivised one would confirm the dominance of the professional logic, whereas a significant difference would disprove it. Whereas controlled experiments are hardly ever possible, comparable circumstances sometimes occur naturally. This is the case for example when very similar research projects are carried out by different teams, one working in conditions close to those provided by the treatment, the other not.

The second best design implies thus to observe and record the partial effect of political institutions on research decisions as they appear to the researcher. In the absence of formal 'treatment' – the presence of political institutions cannot be contrasted with their absence – I am left with comparing situations where the effect of the performative logic is relatively stronger or weaker. Such variations have in the past provided an interesting setting for the study of the relations between democratic institutions and the enforcement of air pollution legislation in the US [90].

Study sponsorship

The first strategy consists in identifying the sponsor of the study. Confirmation bias estimated in this way is better known in the literature as 'funding bias' or 'sponsorship bias', however these terms are synonymous [91]. Empirical studies of funding bias have mainly been undertaken in the area of biomedical research, where drug manufacturers, regulators and patient groups often perform similar studies.

The results of this research are rather unambiguous and consistent: research findings are influenced by the logic of the sponsoring organisation. In one study, for example, researchers looked into every trial of psychiatric drugs in four academic journals over a ten-year period, finding 542 trial outcomes in total. Industry sponsors got favourable outcomes for their own drugs 78% of the time, while independently-funded trials only gave a positive result in 48% of cases. Competing drugs put up against the sponsor's drug in a trial were more effective only in 28% of cases [92].

The underlying assumption is that studies funded by organisations that do not have a vested interest in the outcome of the trial show a lower risk of confirmation bias. The credibility of this assumption rests on the idea that regulators and patient groups have no vested interest in the drug, which is highly questionable. Government funding can result in bias if the aim is to minimise the cost of therapy that it pays for. Likewise, patient organisations may want what they see as the newest and best medications made available to their membership [91].

In the case of policy research, this design would involve comparing similar studies, some sponsored by governmental organisations, others sponsored by non-governmental organisations. This is possible provided two conditions are met. The

first condition is that there are enough studies sponsored by non-governmental organisations to warrant statistical power. This is possible in medical research, which is a highly regulated market – and thus subjected to multiple controls and investigations (from regulators and patient groups). A similar design might prove difficult with social interventions for the opposite reason: policy evaluations are rarely conducted outside government. The second condition is that the studies funded by governmental and non-governmental organisations be reasonably comparable, not only in terms of the intervention, but also in terms of scope, design, timing, etc. This could also be challenging.

Commitment to the intervention

The second strategy consists in contrasting interventions to which the experimenter is strongly committed with interventions to which the experimenter is weakly committed. In political economy, the notion of commitment has mainly been applied in relation to central banks and monetary policy, where predictability and stability are key performance indicators [93]. A number of empirical studies exist regarding the effects of monetary policy commitment [94]. The notion has been more rarely applied to the executive branch. Uses have been limited so far to the notion of compliance to international commitments [95] and to issues of fiscal policy. Yet, the inability of governments to tolerate an open outcome and accept genuine uncertainty as stipulated in the idea of experimentation resonates with earlier observations by Campbell (1969) who had noted that governments tend to commit to policy politically and thus find it difficult to be seen at fault.

Going back to the area of medical research, a drug manufacturer could be committed to a drug because it represents a radical new breakthrough in treatment. A drug manufacturer is also more likely to be committed to drugs that are expected to generate high economic returns. Thus, confirmation bias could be estimated by comparing the effect of drugs with different levels of FDA ranking of therapeutic novelty or drugs with different sales prospects [47]. Likewise, FDA review times were found to be decreasing in (a) the wealth of the richest organisation representing the disease treated by the drug; and (b) media coverage given to this disease. These results suggest that ‘political influence over drug approval operates primarily through ‘salience signals’ transmitted by groups and media [41].

In the policy context, the financial cost of an intervention is certainly a factor, however it can be difficult to get the information in a reliable and consistent way. The cost can be measured in terms of political capital. Reforms to which the government is strongly committed would be subject to a higher risk of bias than reforms to which the government is weakly committed.

Independent variables

The rest of this paper is dedicated to the operationalization of the research question. I will start with the independent variable, namely the government’s commitment to a given reform. Government is largely a black box, which means that the strength of policy commitments can only be *estimated*. The following section offers three possible strategies.

Pilots and phased introductions

The first strategy consists in comparing the decisions made in two different research contexts. The first context is that of a *policy pilot*, i.e. an intervention trialled for a limited period on a fraction of the territory on which it is meant to be rolled out. In principle, the probability that a pilot will be rolled out nationally is unknown at the time of its launch and contingent on the results of the evaluation. The second context is that of the *phased introduction of a reform*, i.e. a reform for which the probability to be fully implemented is known (and close to 1), but for which setup is similar to that of a pilot. Like pilots, phased introductions – which are known in the UK as *pathfinder pilots* – are evaluated on a small scale and over a limited period. They can be evaluated using the exact same designs and methodology. The only difference is in the government's intention and this intention is usually clearly stated.

The extent to which formal pilots and pathfinder pilots are strictly comparable has been debated. Some have mentioned that the term 'pilot' encompassed vastly different projects [42], [97]. According to Ettelt and Mays, pilots can be used for experimentation, for early implementation, for demonstration and for learning how to operationalise a policy [42]. I would argue that these categories are difficult to apply and not mutually exclusive. Indeed, many of the pilots launched by the Labour government tested new ways of delivering social services. It was assumed that the effectiveness of social policy was hampered by inefficient agencies and organisations. In other words, these were public service reforms as much as they were social policy reforms. Furthermore, these categories indistinctly apply to both formal and pathfinder pilots and can be controlled for.

Manifesto pledges

Comparing pilots and phased introductions is a compelling way of estimating policy commitment; however it is not the most frequent type of policy-making. Another, more widely applicable, way of assessing a government's commitment to a specific reform is to check whether it was announced in the ruling party's manifesto for the previous election [98]. Manifestos may not be something voters care about, however the media have specialised in this activity. The multiplication of 'pledge trackers' (such as *The Guardian*'s) shows how crucial it is for a party to implement its pledges once in government. Against this background, it is unlikely that an office-seeking government will commit large resources to pilot a measure that contributed to its electoral success.

Previous research on electoral pledges finds that politicians fulfil most of their electoral promises when they are in power. Pomper and Lederman (1980) find that from 1944–1976, 79% of the pledges proposed by the winning party in the US were fulfilled. Rallings (1987) concludes that 64 of the British pledges from 1945 to 1979 were implemented. Royed (1996) studied British and US electoral pledges during the 1980s. She finds that the British Conservative party implemented more than 80% of its electoral pledges while in government. She also studied British parties in opposition and finds that they obtain much lower fulfilment rates. Only 15% and 32% of the pledges proposed by the Labour party in Britain in 1979 and 1983,

respectively, were fulfilled. Royed (1996) also studied the US case during the 1980 and 1984 electoral cycles and found that even though the Democrats had a majority of seats in at least one of the houses during these years, the Reagan administration was able to act upon 60% of its electoral pledges. Artes (2013) and Chaney (2013) have also contributed to this literature.

In looking at the effect that parties have on policy, manifestos offer a good prediction of what parties will do when in office. This claim is supported by Klingemann, Hofferbert and Budge (1994), who find that government party programmes are remarkably well reflected in post-election priorities, measured as percentages of central government spending in major areas, that is to say that their expenditure reflects the differential issue saliency written into their party manifestos. Based on data from 1970-1979, Rose found that contrary to popular belief governments implemented a large proportion of their manifesto pledges, noting that Labour governments 'acted upon' 55% of their manifesto pledges whilst Conservative governments 'acted upon' 80% of theirs.

Seniority of the 'reform champion'

The problem with manifesto pledges is that they are better suited for studies comparing a large number of heterogeneous policies.

An alternative consists in identifying the 'champion' of the reform. Policy reforms are often introduced by a member of the government; and the seniority of the endorser can be taken as an indication of the salience of the reform. The announcement of a reform can be seen as a delegation issue, whereby each principal, from the Prime minister to the mid-level bureaucrat can decide whether to be the 'manager', taking direct responsibility for the outcome, or the 'chair of the board' overlooking operations [40]. Given politicians' propensity to avoid blame even when that implies not getting credit [39], [40], a reform announced by the Prime minister can be considered more salient than a reform announced by a senior minister. Likewise, a reform announced by a senior minister is considered more salient than a reform introduced by a junior minister. It follows from this that the reforms for which no public announcement is made (which might occur when the reform can be implemented through secondary legislation or statutes) are the least salient.

Using the champion of the reform as independent variable has an additional benefit: it allows more contrast and thus greater measurement validity than dichotomous variables on phased introductions and manifesto pledges. Indeed, there is no reason to believe that such a commitment be so clear-cut. Ordinal variables measuring government policy preferences have been used in studies analysing governments' responsiveness to public opinion [105].

Dependent variables

The following section is concerned with the research decisions that are most likely to reflect confirmation bias. According to Sackett (1979), confirmation bias can occur in any one of seven stages of the research cycle, from the formulation of the research

question to the reporting of findings. Three of these stages are reviewed below as they offer an interesting window for the study of confirmation bias.

Research duration

The first such window is the duration of the research project. Time is a precious resource for both the researcher and the policymaker, but for opposite reasons.

On the one hand, researchers committed to scientific norms will often push for longer research projects. First, repeated measurements are recommended to reduce the statistical phenomenon of regression to the mean which happens when unusually large or small measurements are followed by measurements that are closer to the mean [106], [107]. Second, the psychological literature shows that individuals display different behaviours to novel and usual signals, for example in an experimental context [108]. Thus an individual's response to a new policy is likely to be different as he gets used to it. Third, setting up a research project often takes time; especially when it involves the training of the policy implementers. Researchers eager to generate high-quality evidence are thus more likely to report on the long-term effect of the treatment.

On the other hand, a commitment to the intervention will lead to speedier research. Waiting is politically costly. People suffering from unemployment or crime want a solution to their problems, and incentivised policy-makers want to deliver it as early as possible. There is ample evidence that politicians and business leaders are hard-pressed to deliver before the next election or generate a rapid return on investment. Research within the finance and accounting literatures finds that managers do sacrifice (at least some) long-term investments in response to pressure from the capital markets (Graham et al's (2005)). Similarly Bartov (1993), Bushee (1998), Dechow & Sloan (1991), and Penman & Zjang (2002), all report evidence consistent with the idea that managers sell assets, cut R&D or reduce earnings to meet earnings targets. More recently Benner (2007, 2010) has suggested that firms going through significant technological transitions face particularly intense pressure, causing them to reduce capital investment and investment in R&D [116].

In light of the above, I argue that, other things being equal, shorter research projects denote confirmation bias. The literature on industry-sponsored clinical trials provides evidence to support this claim [41], [47], [117]. A 2010 review compared around a hundred truncated clinical trials and four hundred matched trials that ran their natural course to the end: the truncated trials reported much bigger benefits, overstating the usefulness of the treatments they were testing by about a quarter [118] [119], [120]. Evidence from the policy area is thinner but highlights a similar phenomenon. Anecdotes and interviews have concurred to stress that the greatest source of incompatibility between research and policy rested on the conflict between their respective cycles [97], [121]–[123].

Sampling decisions

Another important decision likely to be affected by confirmation bias is the sampling of the units who will be part of the study.

From a scientific viewpoint, this decision is dominated by the need to have a sample that is as representative of the population at large as possible. A study which conclusions hold over variations in persons, settings, treatments and outcomes is said to have *external validity*. The method most often recommended for achieving this close fit is the use of formal probability sampling [124], [125]. Regardless of the method used, sampling is usually seen as a difficult decision to make, with uncertain results. Using the examples of the *Bangladesh Integrated Nutrition Project* and the *California Class-Size Reduction Program* which both failed in replicating effective interventions evaluated with RCTs, [126] have argued that experiments could not alone support the expectation that a policy will work outside the testbed, given the importance of logistical and contextual factors in the success of a social policy. Whether or not a 'sufficient' level of external validity can be achieved, it is safe to say that researchers committed to the scientific logic will choose their samples in a way that guarantees generalizability.

A sample put together with a view to favour the intervention is expected to be *exemplary* rather than random, heterogeneous or typical. The first reason has to do with the fact that research in a political or market context is skewed towards application. Thus, when the research is carried out *ex post*, it can be tempting for the principal to focus the evaluation on the individuals or groups who seem to have better responded to the intervention. When the research is carried out *ex ante*, the principal may also have an interest in testing the intervention on atypical individuals or groups, for example with a view to increase the probability of generating flattering results. The other reason is basic risk aversion. Research has shown that politicians are motivated primarily by the desire to avoid blame rather than by seeking to claim credit for their decisions [39].

Against this background, I claim that, other things being equal, the representativeness of research samples can be seen as a test for the relation between the researcher and the policy-maker. Here again, the literature on clinical trials suggests that this claim is not unfounded. There is a fairly large – and growing – number of studies pointing to the flimsiness of medical trials' external validity [127]–[131]. For example, one such study took 179 representative asthma patients from the general population and looked at how many would have been eligible to participate in a selection of asthma treatment trials [132]. The answer was 6% on average. Flimsy external validity means that a trial is irrelevant to real-world populations.

Outcome reporting

The third decision that is most likely to reflect confirmation bias is the reporting of evaluation outcomes.

A researcher committed to scientific norms is expected to report findings *in full*, according to pre-specified research questions, theories and variables. Specifying the method from the outset of the research process means that outcomes cannot be manipulated, for example, in order to present positive outcomes. Therefore, provided they apply similar methods, different researchers are likely to report the same results, whether these results are positive, negative or null. The recent years have witnessed the multiplication of initiatives meant to standardise reporting such as CONSORT or COMET.

Conversely, a researcher committed to the intervention is expected to report outcomes *selectively*. Research findings are anything but neutral. In highly regulated industries such as pharmaceutics, an inconclusive trial means that a new drug will not be approved by the regulator. The medical literature highlights a number of recurrent strategies to present these findings in accordance with the interest of the principal. One of them consists in measuring uninformative surrogate outcomes (such as blood pressure or cholesterol rather than the prevalence of specific events such as heart attack or death) or in changing the outcome once the trial is finished [133]–[135]. Another strategy consists in bundling outcomes in a way that changes the presentation of results from negative to positive or from insignificant to significant, for example through the use of composite health indicators [136], [137]. A third strategy implies ignoring the drop-outs that inevitably occur during a trial, which can result in dramatically overstating the benefit of a treatment [138].

As with the two previous outcomes discussed in this paper, evidence of the effect of the performative logic on the reporting of outcomes is scarce. Although there is no evidence that a stronger performative logic is positively correlated with a more selective reporting of outcomes, the literature suggests that it will create pressure on evaluators. A recent web survey of some 200 academics having done policy research for the British government since 2005 indicates that government officials were more likely to propose changes affecting the interpretation of findings or their weight than not. However, it is less clear from the survey whether the requested changes did help produce supportive reports [30]. Beyond this survey, the evidence base consists mostly of some anecdotes, such as Metcalf's report of the pressure exerted by the US Department of Agriculture during the evaluation of the National School Lunch Program [139].

Relevance of the analogy between medical research and policy evaluation

Throughout this paper, I have shown that the literature on the influence of industry sponsorship on the scientific credibility of clinical trials could serve as a useful guide for the study of policy evaluations. Indeed, this literature has very effectively analysed the conundrum which researchers face when they become agents. However, the comparison has also important limitations, which need to be fully understood before we move on to the empirical part of this work. The following section briefly discusses the similarities and differences between medical research and social policy research.

The main similarity between clinical trials and social policy evaluation is that they are both a type of *applied research*. In other words, neither is conducted with the primary purpose of advancing knowledge. Rather, they are meant to inform important decisions about the development of a product or policy, which the organisation is already committed to launch (to varying degrees). This change makes the conduct of research somehow more complex. Each decision not only needs to satisfy the norms of science, it also needs to support the aims of the organisation. Thus, both clinicians and policy researchers have to find the right balance between *professionalism* and *loyalty* to their employer – or reciprocation of 'favours' in the case of contract research.

However, there are also important differences between these two types of research. The first difference is in their purpose. Clinical research is essentially *confirmatory*, i.e. it quantifies the extent to which deviations from a model could be expected to occur by chance. This is due to the fact that (i) health-related variables are easily quantifiable and (ii) medical treatments entail a risk. Medical treatments can not only fail to cure life-threatening diseases, they can also create other diseases and even kill. This is why new drugs have to undergo a series of four consecutive clinical trials[2], all using randomised controlled trials (RCTs). It is commonly admitted in the medical community that RCTs are the most robust way of evaluating the efficacy of a new treatment.

Conversely, social policy research is essentially *exploratory*, i.e. meant to isolate patterns and features of the data. There is no restriction regarding the type of research that can be used to evaluate the effect of a social intervention. Impact evaluations can be conducted using any kind of design (experimental, quasi-experimental or non-experimental) and any kind of data. Indeed, the idea of a 'gold standard' in social policy evaluation is a highly contested one. Thus, social policy reforms are frequently rolled out based on evidence that the intervention was properly implemented, or that beneficiaries were satisfied with the intervention. This 'flexible' approach to research means that there is little consistency across studies in terms of research questions, data and design. Furthermore, the absence of research protocols means that the risk of spin is high.

The second difference is that in medical research, evidence of the cost-effectiveness of a drug has an instrumental use: it is the single most important piece of information that will be considered by regulators in their decision to authorise the drug. When such information is clear and unambiguous, the approval process can be relatively straightforward. In contrast, policy evaluation results have a more conceptual use: governments are free to use results as they see fit and are by no means bound to the conclusions and recommendations of evaluators. Other, non-scientific considerations play an equally, and perhaps greater, role in shaping social policy. Those include ethics, morality, legality, policy commitments and political support. For all these matters, the 'expert' is the elected politician, not the scientist. Unlike medicines and healthcare products regulators, social policy-makers can legitimately discard evaluation results that are found unacceptable or undesirable. Importantly, such a decision implies that an evaluation was conducted in the first place. This is a strong assumption given that no government in the world is subject to a formal obligation to evaluate social interventions.

Financial stakes are a third, major difference. The development process from patent filing to product launch has been estimated to take an average of 12 years at a total cost of some £200 million. In contrast, the costs related to the development of social interventions seem to be much lower. For example, a 2008 Report from the NAO found that, between 2002 and 2006, the DWP had spent about £40 million on initiatives targeted specifically at ethnic minority employment. These included the Ethnic Minority Outreach pilot (£31 million spent between April 2002 and September 2006), the Ethnic Minority Flexible Fund (£6.8 million spent between April 2004 and March 2006) and the Specialist Employment Advisers pilot (£1.5 million spent in 2004-2006). Other initiatives were trialled for a fraction of these costs (for example,

[2] Phase 1: Screening for safety; Phase 2: testing the efficacy of the drug, usually against a placebo; Phase 3: confirmatory study; Phase 4: post-marketing studies delineating additional information on the drug's benefits, risks and optimal uses.

the Mental Health Court pilot and the Virtual Court Pilots were both implemented by the Ministry of Justice for an average cost of £400,000).

Conclusion

This paper was set out to propose a theoretical framework for the study of the influence of political institutions on policy research decisions.

The first objective of this paper was to identify the strengths and the weaknesses of the existing literature, as well as possible research gaps. The review has shown that, collectively, we know a lot about the influence of institutions on policy research. However, this knowledge is fragmented across disciplines and supported by an excessive number of concepts and theories. This paper was unable to review the literature in a systematic way but identified the most significant contributions in research methods, sociology, public policy and public administration. It recommended carrying out more research to identify the key decision-makers and to formulate a more parsimonious theory that would be applicable to a broad range of countries and policy areas.

The second objective of this paper was to lay the foundations an empirical strategy for future research in this area. The notion of confirmation bias, used in research methods and social psychology to qualify the tendency of individuals to favour information that supports prior beliefs and hypotheses, emerged as the most desirable option. There is an abundant literature looking at confirmation bias at the individual level and at the organisational level, particularly in medical research. Thus, the question of the effect of political institutions on policy research would be most effectively addressed by questioning the prevalence of confirmation bias in government-funded research. This entails a two-step approach. First, the scientific credibility of the research decisions made by the relevant government(s) must be systematically investigated, based on a number of common research prescriptions. Second, the effect if policy commitments on the scientific credibility of these research decisions must be assessed.

There are many questions that remained unanswered. I will mention three. First, to the extent that political institutions do influence research, this influence must be context-specific. So we need to understand the contexts in which the effect of political institutions is relatively stronger/weaker. Second, we need to know what happens at the individual level. In particular, we need to understand who the actors are and the type of incentives they are subject to. I have shown that the public policy and the public administration literatures disagree on that point; however this could also be due to the fact that they tend to investigate separate policy areas. More detailed accounts in this area would help researchers make credible assumptions regarding decision-making processes.

Bibliography

[1] E. Schattschneider, *The Semisovereign People: A Realist's View of Democracy in America*. Holt, Rinehart and Winston, 1960.

[2] S. Ehrlich, *Access Points: An Institutional Theory of Policy Bias and Policy Complexity*. New York: Oxford University Press, 2011.

[3] S. Binder, R. Rhodes, and B. Rockman, *The Oxford Handbook of Political Institutions*. Oxford University Press, 2008.

[4] G. Peters, *Institutional Theory in Political Science: The New Institutionalism*, 3rd ed. Continuum Publishing Corporation, 2011.

[5] V. Lowndes and M. Roberts, *Why Institutions Matter: The New Institutionalism in Political Science*. Palgrave Macmillan, 2013.

[6] R. Greenwood, C. Oliver, R. Suddaby, and K. Sahlin-Andersson, *The SAGE Handbook of Organizational Institutionalism*. SAGE, 2008.

[7] M. Bovens, P. 't Hart, and S. Kuipers, "The politics of policy evaluation," in *The Oxford Handbook of Public Policy*, M. Moran, B. Goodin, and M. Rein, Eds. New York: Oxford University Press, 2008, pp. 319–335.

[8] W. Dunn, *Public Policy Analysis: An Introduction*, 3rd ed. Upper Saddle River, NJ: Prentice Hall, 2004.

[9] S. Nagel, *Handbook of Policy Evaluation*. Thousand Oaks, CA: Sage, 2002.

[10] C. Weiss, "Have We Learned Anything New About the Use of Evaluation?," *Am. J. Eval.*, vol. 19, no. 1, pp. 21–33, 1998.

[11] S. Bocking, *Nature's Experts: Science, Politics, and the Environment*. Rutgers University Press, 2004.

[12] R. Merton, "A note on science and democracy," *J. Leg. Polit. Sociol.*, vol. 1942, no. 1, pp. 115–126, 1942.

[13] H. Zuckerman, "The sociology of science ," in *Handbook of sociology*, N. Smelser, Ed. Newbury Park: SAGE Publications, 1988, pp. 511–574.

[14] C. Weiss, "Knowledge creep and Decision Accretion," *Sci. Commun.*, vol. 1, no. 3, pp. 381–404, 1980.

[15] J. Beyer and H. Trice, "The Utilization Process: A Conceptual Framework and Synthesis of Empirical Findings," *Adm. Sci. Q.*, vol. 27, no. 4, pp. 591–622, 1982.

[16] S. Innvaer, G. Vist, and M. Trommald, "Health policy-makers' perceptions of their use of evidence: a systematic review," *J. Health Serv. Res. Policy*, vol. 7, no. 4, pp. 239–244, 2002.

[17] J. Lavis, D. Robertson, J. Woodside, and et al., “How can research organizations more effectively transfer research knowledge to decision makers?,” *Milbank Q.*, vol. 81, no. 2, pp. 171–172, 2003.

[18] A. Ritter and K. Lancaster, “Measuring research influence on drug policy: A case example of two epidemiological monitoring systems,” *Int. J. Drug Policy*, vol. 24, no. 1, pp. 30–37, 2013.

[19] K. Murphy and P. Fafard, “Taking power, politics, and policy problems seriously: the limits of knowledge translation for urban health research,” *J. Urban Heal.*, vol. 89, no. 4, pp. 723–732, 2012.

[20] A. Luke, “Generalizing across borders policy and the limits of educational science,” *Educ. Res.*, vol. 40, no. 8, pp. 367–377, 2011.

[21] N. Amara, M. Ouimet, and R. Landry, “New Evidence on Instrumental, Conceptual, and Symbolic Utilization of University Research in Government Agencies,” *Sci. Commun.*, vol. 26, no. 1, pp. 75–106, 2004.

[22] D. Carpenter, “Is Health Politics Different?,” *Annu. Rev. Polit. Sci.*, vol. 15, pp. 287–311, 2012.

[23] D. Carpenter and G. Krause, “Reputation and Public Administration,” *Public Adm. Rev.*, vol. 72, no. 1, pp. 26–32, 2012.

[24] D. Carpenter, *Reputation and Power: Organizational Image and Pharmaceutical Regulation at the FDA*. Princeton University Press, 2010.

[25] D. Carpenter, *The Forging of Bureaucratic Autonomy: Reputations, Networks and Policy Innovation in Executive Agencies 1862 - 1928*. Princeton: Princeton University Press, 2001.

[26] S. Lavertu and D. Weimer, “Federal Advisory Committees, Policy Expertise, and the Approval of Drugs and Medical Devices at the FDA,” *J. Public Adm. Res. Theory*, vol. 21, no. 2, pp. 211–237, 2011.

[27] C. Hood and D. Jones, *Accident and Design - Contemporary Debates in Risk Management*. London: UCL Press, 1996.

[28] S. Boehmer-Christiansen, “Global climate protection policy: The limits of scientific advice,” *Glob. Environ. Chang.*, vol. 4, no. 2, pp. 140–159, 1994.

[29] M. Feldman and J. March, “Information in Organizations as Signal and Symbol,” *Adm. Sci. Q.*, vol. 26, no. 2, pp. 171–186, 1981.

[30] The LSE GV314 Group, “Evaluation under contract. Government pressure and the production of policy research,” *Public Adm.*, vol. 92, no. 1, pp. 224–239, 2014.

[31] D. Breslau, “The Political Power of Research Methods: Knowledge Regimes in U. S. Labor-Market Policy,” *Theory Soc.*, vol. 26, no. 6, pp. 869–902, 1997.

[32] E. Brodkin and A. Kaufman, “Policy Experiments and Poverty Politics,” *Soc. Serv. Rev.*, vol. 74, no. 4, pp. 507–532, 2000.

[33] R. Rogers-Dillon, *The Welfare Experiments: Politics and Policy Evaluation* . Stanford: Stanford University Press, 2004.

[34] M. Tonry, *Punishment and Politics: Evidence and Emulation in the Making of English Crime Control Policy*. London: Willan , 2004.

[35] C. Lindblom, “The Science of ‘Muddling Through,’” *Public Adm. Rev.*, vol. 19, no. 2, pp. 79–88, 1959.

[36] M. Monaghan, “The Complexity of Evidence: Reflections on Research Utilisation in a Heavily Politicised Policy Area,” *Soc. Policy Soc.*, vol. 9, no. 1, pp. 1–12, 2010.

[37] J. Henig, “Politicization of Evidence: Lessons for an Informed Democracy,” *Educ. Policy*, vol. 23, no. 1, pp. 137–160, 2009.

[38] J. Henig, *Spin Cycle: How Research Is Used in Policy Debates, The Case of Charter Schools*. Russell Sage Foundation/The Century Foundation, 2008.

[39] R. Weaver, “The Politics of Blame Avoidance,” *J. Public Policy*, vol. 6, no. 4, pp. 371–398, 1986.

[40] C. Hood, *The Blame Game: Spin, Bureaucracy and Self-Preservation in Government*. Princeton and Oxford: Princeton University Press, 2011.

[41] D. Carpenter, “Groups, the Media, Agency Waiting Costs, and FDA Drug Approval,” *Am. J. Pol. Sci.*, vol. 46, no. 3, pp. 490–505, 2002.

[42] S. Ettelt and N. Mays, “Health policy piloting in England in the 2000s: has the drive towards ‘evidence based policy’ resolved long-standing dilemmas in relation to the purpose of piloting?,” in *Policy Pilots and Evaluation. Health History and Policy Seminar*, 2013.

[43] M. Weber, “Science as a Vocation,” in *Essays in Sociology*, H. Gerth and W. Mills, Eds. New York: Oxford University Press, 1946, pp. 129–156.

[44] J. MacDonald and W. Franko, “Bureaucratic Capacity and Bureaucratic Discretion: Does Congress Tie Policy Authority to Performance?,” *Am. Polit. Res.* , vol. 35, no. 6, pp. 790–807, 2007.

[45] J. MacDonald, “Limitation Riders and Congressional Influence over Bureaucratic Policy Decisions,” *Am. Polit. Sci. Rev.*, vol. 104, no. 04, pp. 766–782, 2010.

[46] D. Carpenter, “Protection without capture: Product approval by a politically responsive, learning regulator ,” *Am. Polit. Sci. Rev.*, vol. 98, no. 4, pp. 613–631, 2004.

[47] D. Dranove and D. Meltzer, “Do Important Drugs Reach the Market Sooner?,” *RAND J. Econ.*, vol. 25, no. 3, pp. 402–423, 1994.

[48] G. Krause and J. Douglas, “Institutional Design versus Reputational Explanations for Agency Performance: Evidence from U.S. Government Macroeconomic and Fiscal Projections,” *J. Public Adm. Res. Theory*, vol. 15, no. 2, pp. 281–306, 2005.

[49] N. Steneck, “The role of professional societies in promoting integrity in research ,” *Am. J. Heal. Behav.* , vol. 27, pp. 239–247, 2003.

[50] L. Andersen, “Professional norms, public service motivation and economic incentives: What motivates public employees?,” 2007.

[51] E. Ostrom, “An Agenda for the Study of Institutions,” *Public Choice*, vol. 48, pp. 3–25, 1986.

[52] I. Mitroff, “Norms and Counter-Norms in a Select Group of the Apollo Moon Scientists: A Case Study of the Ambivalence of Scientists,” *Am. Sociol. Rev.*, vol. 39, no. 4, pp. 579–595, 1974.

[53] D. Hull, *Science as a Process: An Evolutionary Account of the Social and Conceptual Development of Science*. University of Chicago Press, 1988.

[54] T. Pinch, *Confronting Nature: The Sociology of Solar-Neutrino Detection*. Dordrecht: Dordrecht Reidel, 1986.

[55] J. Sapp, *Where the Truth Lies: Franz Moewus and the Origins of Molecular Biology*. New York: Cambridge University Press, 1990.

[56] M. Mulkay, “Norms and ideology in science,” *Soc. Sci. Inf.*, vol. 15, pp. 637–656, 1976.

[57] D. Sackett, “Bias in analytic research,” *J Chronic Dis.*, vol. 32, no. 1–2, pp. 51–63, 1979.

[58] M. Gale and L. Ball, “Does Positivity Bias Explain Patterns of Performance on Wason’s 2-4-6 task?,” in *Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society*, W. Gray and C. Schunn, Eds. Routledge, 2002, p. 340.

[59] S. Plous, *The Psychology of Judgment and Decision Making*. New York: McGraw-Hill, 1993.

[60] M. Oswald and S. Grosjean, “Confirmation Bias,” in *Cognitive Illusions: A Handbook on Fallacies and Biases in Thinking, Judgement and Memory*, R. Pohl, Ed. Hove, UK: Psychology Press, 2004, pp. 79–96.

[61] A. Hergovich, R. Schott, and C. Burger, “Biased evaluation of abstracts depending on topic and conclusion: Further evidence of a confirmation bias within scientific psychology. ,” *Curr. Psychol.*, vol. 29, no. 3, pp. 188–209, 2010.

[62] R. Cialdini, J. Cacioppo, R. Bassett, and J. Miller, “Low-ball procedure for producing compliance: Commitment then cost.,” *J. Pers. Soc. Psychol.*, vol. 36, no. 5, pp. 463–476, 1978.

[63] P. Tetlock, *Expert Political Judgment: How Good Is It? How Can We Know?* Princeton: Princeton University Press, 2005.

[64] R. Nickerson, “Confirmation bias: A ubiquitous phenomenon in many guises,” *Rev. Gen. Psychol.*, vol. 2, no. 2, pp. 175–220, 1998.

[65] J. Koehler, “The influence of prior beliefs on scientific judgments of evidence quality,” *Organ. Behav. Hum. Decis. Process.*, vol. 56, pp. 28–55, 1993.

[66] M. Mahoney, “Publication prejudices: An experimental study of confirmatory bias in the peer review system,” *Cognit. Ther. Res.*, vol. 1, no. 2, pp. 161–175, 1977.

[67] R. Cialdini, “Crafting normative messages to protect the environment,” *Curr. Dir. Psychol. Sci.*, vol. 12, no. 4, pp. 105–109, 2003.

[68] E. Fehr and S. Gächter, “Fairness and Retaliation: The Economics of Reciprocity,” *J. Econ. Perspect.*, vol. 14, no. 3, pp. 159–181, 2000.

[69] D. Regan, “Effects of a favor and liking on compliance,” *J. Exp. Soc. Psychol.*, vol. 7, no. 6, pp. 627–639, 1971.

[70] Institute of Medicine, *Conflict of Interest in Medical Research, Education, and Practice*. Washington, DC: The National Academies Press, 2009.

[71] K. Austad, J. Avorn, and A. Kesselheim, “Medical Students’ Exposure to and Attitudes about the Pharmaceutical Industry: A Systematic Review,” *PLOS Med.*, vol. 8, no. 5, 2011.

[72] B. Glaser and L. Bero, “Attitudes of academic and clinical researchers toward financial ties in research: a systematic review,” *Sci. Eng. Ethics*, vol. 11, no. 4, pp. 553–573, 2005.

[73] Z. Kunda, *Social Cognition: Making Sense of People*. MIT Press, 1999.

[74] C. Fine, *A Mind of Its Own: How Your Brain Distorts and Deceives*. Cambridge (UK): Icon Books Ltd, 2006.

[75] D. A. Redelmeier and A. Tversky, “On the belief that arthritis pain is related to the weather.,” *Proc. Natl. Acad. Sci.*, vol. 93, no. 7, pp. 2895–2896, Apr. 1996.

[76] C. Turner and G. Spilich, “Research into smoking or nicotine and human cognitive performance: does the source of funding make a difference?,” *Addiction*, vol. 92, no. 11, pp. 1423–1426, 1997.

[77] Becker-Brüser W, “Research in the pharmaceutical industry cannot be objective (article in German),” *Z Evid Fortbild Qual Gesundhwes*, vol. 104, no. 3, pp. 183–189, 2010.

[78] C. Bruce Baker, M. Johnsrud, M. Crismon, R. Rosenheck, and S. Woods, “Quantitative analysis of sponsorship bias in economic studies of antidepressants,” *Br. J. Psychiatry*, vol. 183, no. 6, pp. 498–506, 2003.

[79] D. Westen, P. Blagov, Harenski, C. Kilts, and S. Hamann, “Neural bases of motivated reasoning: an FMRI study of emotional constraints on partisan political judgment in the 2004 U.S. Presidential election,” *J Cogn Neurosci.*, vol. 18, no. 11, pp. 1947–1958, 2006.

[80] C. Lord, L. Ross, and M. Lepper, “Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence,” *J. Pers. Soc. Psychol.*, vol. 37, no. 11, pp. 2098–2109, 1979.

[81] C. S. Taber and M. Lodge, “Motivated Skepticism in the Evaluation of Political Beliefs,” *Am. J. Pol. Sci.*, vol. 50, no. 3, pp. 755–769, Jul. 2006.

[82] K. Stanovich and R. West, “Natural myside bias is independent of cognitive ability,” *Think. Reason.*, vol. 13, no. 3, pp. 225–247, 2007.

[83] A. Gilinsky and B. Judd, “Working memory and bias in reasoning across the life span,” *Psychol. Aging*, vol. 9, no. 3, pp. 356–371, 1994.

[84] S. Handley, A. Capon, M. Beveridge, I. Dennis, and J. Evans, “Working memory, inhibitory control and the development of children’s reasoning,” *Think. Reason.* , vol. 10, no. 2, pp. 175–195, 2004.

[85] J. Kokis, R. Macpherson, M. Toplak, R. West, and K. Stanovich, “Heuristic and analytic processing: Age trends and associations with cognitive ability and cognitive styles,” *J. Exp. Child Psychol.*, vol. 83, pp. 26–52, 2002.

[86] K. Knorr, “Policy-makers’ use of social science knowledge: symbolic or instrumental?,” in *Using social research in public policy making*, C. Weiss, Ed. Lexington: Lexington Books, 1977, pp. 165–182.

[87] C. Weiss, “The Many Meanings of Research Utilization,” *Public Adm. Rev.*, vol. 39, no. 5, pp. 426–431, 1979.

[88] D. Pelz, “Some Expanded Perspectives on Use of Social Science in Public Policy,” in *Major Social Issues: A Multidisciplinary View*, J. Yinger and S. Cutler, Eds. New York: The Free Press, 1978, pp. 346–357.

[89] E. Fama, “Agency Problems and the Theory of the Firm ,” *J. Polit. Econ.*, vol. 88, no. 2, pp. 288–307, 1980.

[90] B. Wood, “Principals, Bureaucrats, and Responsiveness in Clean Air Enforcements,” *Am. Polit. Sci. Rev.*, vol. 82, no. 1, pp. 213–234, 1988.

[91] J. Lexchin, “Sponsorship bias in clinical research,” *Int J Risk Saf Med.*, vol. 24, no. 4, pp. 233–242, 2012.

[92] R. Kelly, L. Cohen, R. Semple, and et al., “Relationship between drug company funding and outcomes of clinical psychiatric research ,” *Psychol. Med.*, vol. 36, no. 11, pp. 1647–1656, 2006.

[93] Y. Nakazono and K. Ueda, “Policy commitment and market expectations: Lessons learned from survey based evidence under Japan’s quantitative easing policy,” *Japan World Econ.*, vol. 25–26, pp. 102–113, Jan. 2013.

[94] N. Baba, S. Nishioka, N. Oda, M. Shirakawa, K. Ueda, and H. Ugai, “Japan’s Deflation, Problems in the Financial System, and Monetary Policy,” *Monet. Econ. Stud.*, vol. 23, no. 1, pp. 47–111, 2005.

[95] J. Kelley, “Who Keeps International Commitments and Why? The International Criminal Court and Bilateral Nonsurrender Agreements,” *Am. Polit. Sci. Rev.*, vol. 101, no. 03, p. 573, Jul. 2007.

[96] D. Campbell, “Reforms as experiments,” *Am. Psychol.*, vol. 24, no. 4, pp. 409–429, 1969.

[97] R. Jowell, “Trying It Out. The Role of ‘Pilots’ in Policy-Making.,” London, 2003.

[98] R. Rose, *Do Parties Make a Difference?* London: Macmillan, 1980.

[99] G. Pomper and S. Lederman, *Elections in America: Control and Influence in Democratic Politics.* . London/New York: Longman, 1980.

[100] C. Rallings, “The Influence of Election Programs: Britain and Canada 1945-1979,” in *Ideology, Strategy and Party Change*, I. Budge, D. Robertson, and D. Hearl, Eds. Cambridge: Cambridge University Press, 1987, pp. 1–14.

[101] T. Royed, “Testing the mandate model in Britain and the United States: Evidence from the Reagan and Thatcher eras,” *Br. J. Polit. Sci.*, vol. 26, no. 1, pp. 45–80, 1996.

[102] J. Artes, “Do Spanish politicians keep their promises?,” *Party Polit.*, vol. 19, no. 1, pp. 143–158, 2013.

[103] P. Chaney, “Electoral discourse analysis of state foreign policy development: exploring the party politicization of the Commonwealth in UK Westminster elections,” *Contemp. Polit.*, vol. 19, no. 2, pp. 203–220, 2013.

[104] H. Klingemann, R. Hofferbert, and I. Budge, *Parties, Policies and Democracy*. Westview Press, 1994.

[105] S. B. Hobolt and R. Klemmemsen, “Responsive Government? Public Opinion and Government Policy Preferences in Britain and Denmark,” *Polit. Stud.*, vol. 53, no. 2, pp. 379–402, Jun. 2005.

[106] A. Barnett, J. Van der Pols, and A. Dobson, “Regression to the mean: what it is and how to deal with it,” *Int. J. Epidemiol.*, vol. 34, no. 1, pp. 215–220, 2005.

[107] S. Stiegler, “Regression towards the mean, historically considered,” *Stat. Methods Med. Res.*, vol. 6, no. 2, pp. 103–114, 1997.

[108] R. Gillespie, *Manufacturing knowledge : a history of the Hawthorne experiments*. Cambridge: Cambridge University Press, 1991.

[109] J. Graham, C. Harvey, and S. Rajgopal, “The economic implications of corporate financial reporting,” *J. Account. Econ.*, vol. 40, no. 1, pp. 3–73, 2005.

[110] P. Dechow and R. Sloan, “Executive incentives and the horizon problem: An empirical investigation,” *J. Account. Econ.*, vol. 14, no. 1, pp. 51–89, 1991.

[111] E. Bartov, “The Timing of Asset Sales and Earnings Manipulation,” *Account. Rev.*, vol. 68, no. 4, pp. 840–855, 1993.

[112] B. Bushee, “The Influence of Institutional Investors on Myopic R&D Investment Behaviour,” *Account. Rev.*, vol. 73, no. 3, pp. 305–333, 1998.

[113] S. Penman and X. Zjang, “Accounting Conservatism, the Quality of Earnings, and Stock Returns,” *Account. Rev.*, vol. 77, no. 2, pp. 237–264, 2002.

[114] M. Benner, “The incumbent discount: stock market categories and response to radical technological change,” *Acad. Manag. Rev.*, vol. 32, no. 3, pp. 703–720, 2007.

[115] M. Benner, “Securities Analysts and Incumbent Response to Radical Technological Change: Evidence from Digital Photography and Internet Telephony,” *Organ. Sci.*, vol. 21, no. 1, pp. 42–46, 2010.

[116] N. Repenning and R. Henderson, “Making the Numbers? ‘Short Termism’ & The Puzzle of Only Occasional Disaster ,” 11-033, 2010.

[117] M. Olson, “Firm Characteristics and the Speed of FDA Approval,” *J. Econ. Manag. Strateg.*, vol. 6, no. 1, pp. 377–401, 1997.

[118] D. Bassler, M. Briel, V. Montori, and et al., “Stopping randomized trials early for benefit and estimation of treatment effects: systematic review and meta-regression analysis.,” *J. Am. Med. Assoc.*, vol. 303, no. 12, pp. 1180–1187, 2010.

[119] V. Montori, P. Devereaux, N. Adhikari, and et al., “Randomized trials stopped early for benefit: a systematic review,” *J. Am. Med. Assoc.*, vol. 294, no. 17, pp. 2203–2209, 2005.

[120] F. Trotta, G. Apolone, Garattini S, and G. Tafuri, “Stopping a trial early in oncology: for patients or for industry?,” *Ann. Oncol.*, vol. 19, no. 7, pp. 1347–1353, 2008.

[121] J. Coleman, “The Use of Social Science in the Development of Public Policy,” *Urban Rev.*, vol. 10, pp. 197–202, 1978.

[122] I. Boa, P. Johnson, and S. King, “The impact of research on the policy process,” London, 82, 2010.

[123] M. Hallsworth, S. Parker, and J. Rutter, *Policy making in the real world*. London: Institute for Government, 2011.

[124] W. Shadish, T. Cook, and D. Campbell, *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin, 2002.

[125] P. Rossi, J. Wright, and A. Anderson, *Handbook of Survey Research*. Orlando and London: Academic Press, 1983.

[126] N. Cartwright and J. Hardie, *Evidence-Based Policy: A Practical Guide to Doing It Better*. New York: Oxford University Press, 2012.

[127] P. Rothwell, “External validity of randomised controlled trials: ‘to whom do the results of this trial apply?,’” *Lancet*, vol. 365, no. 9453, pp. 82–93, 2005.

[128] C. Pratt and L. Moyé, “The cardiac arrhythmia suppression trial: Casting suppression in a different light.,” *Circulation*, vol. 91, no. 1, pp. 245–247, 1995.

[129] M. Zimmerman, I. Chelminski, and M. Posternak, “An illustration of how a self-report diagnostic screening scale could improve the internal validity of antidepressant efficacy trials,” *J. Affect. Disord.*, vol. 80, no. 1, pp. 79–85, 2004.

[130] G. Keitner, M. Posternak, and C. Ryan, “How many subjects with major depressive disorder meet eligibility requirements of an antidepressant efficacy trial?,” *J. Clin. Psychiatry*, vol. 64, no. 9, pp. 1091–1093, 2003.

[131] T.-P. van Staa, H. Leufkens, Zhang B, and et al., “A Comparison of Cost Effectiveness Using Data from Randomized Trials or Actual Clinical Practice: Selective Cox-2 Inhibitors as an Example,” *PLOS Med.*, vol. 6, no. 12, pp. 1–9, 2009.

[132] J. Travers, S. Marsh, M. Williams, and et al., “External validity of randomised controlled trials in asthma: to whom do the results of the trials apply?,” *Thorax*, vol. 62, no. 3, pp. 219–223, 2007.

[133] S. Vedula, L. Bero, and R. Scherer, “Outcome Reporting in Industry-Sponsored Trials of Gabapentin for Off-Label Use,” *N. Engl. J. Med.*, vol. 361, no. 20, pp. 1963–1971, 2009.

[134] A. Chan, A. Hróbjartsson, M. Haahr, and et al., “Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles,” *J. Am. Med. Assoc.*, vol. 291, no. 20, pp. 2457–2465, 2004.

[135] J. Jureidini, L. McHenry, and P. Mansfield, “Clinical trials and drug promotion: Selective reporting of study 329,” *Int. J. Risk Saf. Med.*, vol. 20, no. 1–2, pp. 73–81, 2008.

[136] V. Montori, R. Jaeschke, H. Schünemann, and et al., “User’s guide to detecting misleading claims in clinical research reports,” *Br. Med. J.*, vol. 329, no. 7474, pp. 1093–1096, 2004.

[137] A. Shaughnessy, “What happened to the valid POEMs? A survey of review articles on the treatment of type 2 diabetes,” *Br. Med. J.*, vol. 327, no. 7409, p. 266, 2003.

[138] H. Melander, J. Ahlqvist-Rastad, G. Meijer, and et al., “Evidence b(i)ased medicine--selective reporting from studies sponsored by pharmaceutical industry: review of studies in new drug applications,” *Br. Med. J.*, vol. 326, no. 7400, pp. 1171–1173, 2003.

[139] C. Metcalf, “Threats to independence and objectivity of government-supported evaluation and policy research,” *J. Policy Anal. Manag.*, vol. 27, no. 4, pp. 927–934, 2008.