Performative Power in Science[1]

Philippe van Basshuysen

Centre for Ethics and Law in the Life Sciences, Leibniz University Hannover

philippe.van.basshuysen@cells.uni-hannover.de

Abstract. Performativity is the capacity of scientific representations (such as models, theories, predictions, or classifications) to alter the phenomena they are supposed to represent. Because one and the same representation may alter its target a lot, a little, or not at all, I argue that we should conceptualize performativity as a function of a representation and its *performative power*, which depends on its reach, acceptance, and its relevance to people. Using this framework, I then argue for a re-evaluation of performativity. Because performative effects can impair scientists' ability to model, classify, explain, or predict (e.g. by steering outcomes away from those predicted) and they raise concerns about the legitimacy of science influencing the social world, performativity is often viewed as a threat to science. In contrast, I argue that we shouldn't be worried about performativity as such, but rather, about *concentrated performative power*, that is, that the representations issued by individual scientists, or groups of scientists with uniform views, gain a power that is not in line with their epistemic credentials. To eradicate such power concentrations and to secure science's proper role in a democratic society, we may sometimes hope for more, rather than fewer, performative effects.

Keywords. Performativity; power; monopoly; democracy; values

1. Introduction

A scientific model, prediction, explanation, or classification is *performative* when it interacts with the phenomenon it represents -- by altering it or even bringing it about. Performativity has long been recognized, especially in the social sciences: for instance, a prediction of a bank failure may induce customers to withdraw their money all at once, thus making the prediction become a self-fulfilling prophecy (Merton 1948). Yet, the same type of prediction can differ greatly in its performative force: *my* prediction that a bank will collapse tomorrow will very likely not kick off any tangible effects; whereas the same type of prediction may have significant effects if it is made, for instance, by an influential economic think tank, backed by prominent economists, and is widely disseminated through the news and social media. This basic observation, I will argue, has significant implications for how we should conceptualize performativity, and for how performative science should be evaluated and managed.

Concerning conceptualization, I propose that any performative effect can be analyzed as a function of a scientific representation and its *performative power*. Performative power is the representation's ability to influence the course of events, which will crucially depend on its relevance to people, and its reach and acceptance (Morgenstern 1928, 99) – on *how many* and *who* will become acquainted with tokens of the representation, and on whether and to what extent those reached will come to believe in and act upon it. My hypothetical prediction of a

[1] I thank Matthias Ackermann, Markus Ahlers, Atocha Aliseda, George Barimah, Lukas Beck, Mathias Frisch, Catherine Herfeld, Nils Hoppe, Donal Khosrowi, Simon Lohse, Jakob Ortmann, Tanja Rechnitzer, Nadia Ruiz, Jacob Stegenga, and Jannik Zeiser for helpful discussions and/or feedback, as well as the audiences of the International Symposium Ethics of the Sciences (March 2024), the GRK Research Colloquium at Leibniz University Hannover (June 2024), and the Seminario de Investigadores at the Instituto de Investigaciones Filosoficas, UNAM (February 2025).

bank failure, for instance, would have low performative power because not many will become acquainted with it, and of those who do, few will believe it; whereas, uttered by influential and credible economic experts, the same type of prediction might be quite powerful. The first contribution of this paper is to provide a framework for reflecting on performative power: answering the questions of what performative power is, how it is generated, and how it might be measured.

As a second contribution, the framework will then be applied to the evaluation of the *legitimacy of performative science*. Ideally, epistemic value and performative power would align: the better the model, the more likely it is to be heard and acted upon. But misalignments are common. Representations with little epistemic merit may still acquire great performative power, resulting in monopolistic influence. Such concentrations of power are epistemically and ethically problematic and, I will argue, should be disrupted. This perspective contrasts with a long tradition that sees performativity as a threat to science's epistemic and ethical integrity. On my account, performativity itself is not the problem. What matters is how performative power is distributed: breaking power concentrations can yield new, sometimes even greater, performative effects – effects we need not regret and may even welcome.

Once we clearly see that the problem is not performativity as such, but rather, the idiosyncratic effects that result from concentrated performative power, the question of how performative science ought to be managed to exclude such power concentrations comes to the fore. The final contribution of this paper is to make progress on this question, by reflecting on the proper place of performative science in a democratic society, and how responsibilities with regard to representations and their performative power ought to be allocated. Scientists, I argue, should be accountable for the epistemic value of their published representations and they should seek to anticipate and forego very detrimental performative effects (cf. Godman and Marchionni 2022). Responsibility for the performative power of scientific results, however, should be distributed more broadly in a democratic society.

Before proceeding, it is important to clarify the scope of the claims advanced in this paper. The framework I develop does not aim to answer the question (on which there is no agreement in the literature to date) of what it means for a representation to be epistemically valuable in a performative world, or to provide a direct operational test for distinguishing epistemically deficient from epistemically valuable representations. Rather, its contribution is conceptual: to make precise how performative effects depend on the interplay between representations and their performative power. Even if we cannot in practice always determine whether a given representation is epistemically deficient, the framework remains useful because it allows us to articulate why concentrated performative power is problematic, where power concentrations arise *given* specific views about the epistemic value of performative representations, and to clarify what is at stake in discussions about managing performative science. In this sense, the framework provides orientation and critical vocabulary to analyze performative science.

The remainder of the paper is structured as follows. Section 2 briefly reviews how performativity has been discussed in the literature. Section 3 develops the conceptual framework, focusing on the notion of performative power. Section 4 applies this framework to argue that the real problem lies not in performativity as such but in concentrated performative power. Section 5 considers how performative science might be managed to avoid such concentrations. Section 6 concludes.

2. What Is Performativity?

Since performativity has been used to describe various phenomena in a range of different fields, this section issues a brief clarification of how this term will be used throughout, as well as a bird's eye view of the relevant literature. In a recent paper, scientific models, theories and predictions are defined as performative whenever their publication or dissemination impacts on their intended target in such a way that the fit with that target is altered as a result (Ortmann 2025, 5). While this definition nicely captures pertinent examples and is suitable for the purposes of this paper, I add a few clarificatory remarks:

First, the kinds of vehicles capable of kicking off performative effects are manifold. They include models, theories, and predictions, but also classifications, explanations, figures and graphs, and other results and aspects of science (cf. MacKenzie 2006, 35). Moreover, what the vehicle(s) of performative effects are in a given case isn't always clear-cut. For instance, while performativity is often attributed to models (e.g. Khosrowi 2023), it is common that a prediction *based on a model* (rather than the model itself) is broadly disseminated and thereby triggers reactions. A typical case would be an economic prediction based on a mathematical model, which, through its dissemination, influences business decisions, which in turn alter economic outcomes. For present purposes, I shall put to a side the question (which is orthogonal to the arguments made in this paper) of what the vehicle of a given effect is and, as a convenient terminological stipulation, I will use *representation* as an umbrella for all those aspects of science that can serve as vehicles for performative effects.[2]

Second, performativity can be brought about through different mechanisms, or "channels of transmission" (Vergara-Fernández et al. 2023, 15). In particular, performative effects can transpire directly through spontaneous behavioral changes (Guala 2007 calls this "genuine performativity" (24)), or more indirectly mediated through institutional design or policy making ("spurious performativity", ibid.).[3]

Third and relatedly, there is a venerable discussion about the question of whether performativity is a feature pertaining exclusively to the social sciences and economics, in which case this might then be used as a demarcation criterion between the social and the natural

[2] As an anonymous reviewer points out, what performative effects might materialize in a given case, and how these effects ought to be evaluated, may depend on the vehicle bringing about these effects. That is, it might matter whether these vehicles are models, predictions, explanations, classifications, or other kinds of representations. Moreover, different kinds of representations can serve different epistemic goals and their performativity might accordingly raise different duties on the part of scientists. I believe these are valid concerns, but they are orthogonal to the argument of this paper whose modest goal is to introduce and raise awareness of concentrated performative power. If this point can be established concerning predictions (or models, or theories…) this would seem sufficient for my argument to go through. I agree, however, that the literature on performativity tends to be sloppy with regard to the question of what the vehicles of performative effects are, and I believe more work should be done drawing relevant distinctions and thinking about possible implications for our evaluation of performative science. The next section will make a start on this task, by introducing an important type-token distinction with respect to performative representations.

[3] Guala also argues that, at least with respect to economics, the two kinds of performativity are tightly interdependent and can hardly be observed separately from one another (2007, 25). I believe this is true in other fields than economics, too. In the COVID-19 case, for instance, we saw that individuals engaged in social distancing even before government-issued directives took effect, reinforcing these behaviors (see van Basshuysen et al. 2021). If it is correct that these routes cannot typically be neatly separated, this implies that engaging in policy advising does not exempt scientists from considering further performative effects of their research, as these effects may nevertheless kick in. I thank an anonymous reviewer for pressing me on this point.

sciences (see Grünbaum 1956; Merton 1949; Popper 1957; Buck 1963; also cf. Hacking 1995; 1999). For the purpose of this paper, we need not take up a stance on this question, though it may be noted that only the genuine form of performativity would seem to be more confined to the social realm (and even here, examples from the biological and health sciences abound), whereas the indirect route is open to other sciences, too.

Finally, there are a range of other, closely related concepts, including *self-fulfilling* (or *self-defeating*) prophecies (Merton 1948), *reflexive predictions* (Buck 1963), and *looping effects* of human kinds (Hacking 1999), which may be seen as special cases of performativity because each involves a change of the fit of a scientific representation with its target that is due to its dissemination. Inversely, the concept of *reactivity* can be seen as even broader than performativity as it aims to capture not only reactions to scientific results but also reactions at the level of data collection (e.g. when experimental subjects alter their behavior in response to being studied), which aren't directly relevant for the purposes of this paper (though see Marchionni et al. 2024 for an excellent overview).

With these conceptual clarifications in mind, we can now clearly see why philosophers and scientists alike have often worried that performativity presents an obstacle to the epistemic goals of science. When the very dissemination of a representation alters its intended target, this seems to render elusive science's task of accurately representing and modeling phenomena. For instance, Oskar Morgenstern, focusing on economic predictions of the business cycle, argued that, when such a prediction becomes public, businesses will adapt to the predicted events, for example, by increasing or decreasing their production. Because their decisions change, this will change the course of the predicted events, thus making the prediction come out false. Of course, it wouldn't help, then, to simply issue a second prediction, because this new prediction would again lead to changes in business decisions that would make this prediction false, and the same would happen for any further predictions, so that we have a potentially infinite series of false predictions (see Morgenstern 1928, 97-98). Even worse, performativity, as we have seen, plagues a broad range of scientific results and aspects and can be brought about through different mechanisms, and performative effects aren't at all rare but are almost *ubiquitous* and *inevitable* in research that concerns people (cf. Hardt and Mendler-Dünner 2023, 24). Performativity is ubiquitous because people respond to incentives, and scientific results that concern them will frequently change their incentives, for instance, by providing information about the risks or benefits associated with certain kinds of behaviors. It is almost inevitable because results will typically be publicly available, and even if, for instance, a researcher wanted to keep a prediction private, it can in principle be replicated by others, and this will then bring about the effects the initial researcher wished to avoid (see Popper 1957, 13). And, finally, for theorists who see performativity as a demarcation criterion between social science and economics on the one hand (where these effects occur), and natural science on the other (where they don't), a further reason to view performativity with suspicion are the successes of natural science and the comparatively "dismal" status of economics and the social sciences. For is this really a coincidence, or is it rather *due to* performativity that the social sciences have proven (predictively, explanatorily, in terms of interventions etc.) so much weaker?

In addition to these epistemic issues (to which we'll come back in Section 4 below), some have additionally worried about seemingly manipulative and anti-democratic uses of performative science. Some of these arguments were made in response to the view that performativity can sometimes be the purpose of a model, which should be taken into account when we evaluate that model. Against this line of argument (as defended by van Basshuysen et

al. 2021), Eric Winsberg and Stephanie Harvard claim that performativity "is never a legitimate purpose for a model" as this "would be a serious threat to democratic decision making" (2022, 4). Their argument is that models should allow decision makers to estimate whether the benefits associated with a an action or policy would outweigh its costs. However, when the model itself influences outcomes, they claim, it would then be impossible to conduct an impartial cost-benefit analysis (van Basshuysen 2023 shows that this doesn't necessarily follow). Rather, scientists would be constructing a model in a way that would bring about certain outcomes rather than others, thus *presupposing* what the desirable outcome is. But, according to Winsberg and Harvard, scientists aren't democratically legitimated to decide on the public's behalf,[4] which is why treating performativity as a model purpose seems to go against the democratic procedure. Thus, this line of work views performativity not only as an epistemic problem but as a serious threat to the ethical integrity of science and its proper standing in a democratic society. We will come back to the epistemic and ethical/social concerns surrounding performativity in Section 4, where I will argue for a re-evaluation of performative science.

Yet, while the assumption that performativity is detrimental to science has been influential, there is an alternative view that is gaining traction, which also emphasizes positive aspects. Focusing on experimental economics and mechanism design, Francesco Guala argued that performativity here serves as an asset for the social scientist, in designing in desirable ways the institutions through which we interact (2007, p. 3; see also Jimenez-Buedo and Guala 2016; van Basshuysen 2023). Guala's is the first in a growing series of articles that seek to identify constructive uses of performativity, by analyzing a variety of cases, from epidemiology (van Basshuysen et al. 2021) to activist research (Koskinen 2022), financial economics (Vergara-Fernández et al. 2023), and climate change mitigation (Ortmann and Veit 2023). My account is broadly in agreement with this recent literature. However, showing that performativity can have constructive roles to play in science doesn't imply that anything goes; rather, we then need to go about demarcating what are legitimate ways of dealing with the phenomenon (as the literature on values in science teaches us; see Holman and Wilholt 2022). To get this project started, I will next introduce a general conceptual framework, which will assist the evaluation and management of performativity in a broad range of cases.

3. Why Are Some Scientific Representations Performative?

I propose to think of any performative effect as a *function of a scientific representation issued and its corresponding performative power*. This needs explanation; in the remainder of this section, I will first provide an intuitive understanding of my proposal, and second, clarify a) what makes for performative power, b) how it is generated and what it should be attributed to, and c) how it might be measured.

Let's start with an intuitive example:

> *A group of economists publishes a white paper predicting a sharp drop in equities of companies in a particular sector. Some stock market traders first pick up the prediction. Deeming it credible, many of them immediately sell their shares of the respective*

[4] Winsberg and Harvard do not explain why they believe scientists aren't democratically legitimated to decide on the public's behalf, or what would legitimize them to do so. Perhaps what they have in mind is a requirement to be democratically elected, but there are many other forms of democratic legitimation than just election, and scientists can plausibly be viewed as political representatives without being elected (see Hilligardt 2023). Section 5 will come back to this discussion.

> *companies, and they also pass the prediction on to clients and other traders. The prediction is also taken up by news outlets and shared on social media. More people come to believe in the crash being imminent and react by selling their shares. The news spreads further, through newspapers, TV, and the internet, and more people follow suit. The prediction, we imagine, ends up being performative: what would otherwise have been a minor drop ends up being a major crash in equities, due to the reactions to the prediction.*

Even though imagined, I take it that the example depicts typical ways in which a prediction can travel and disperse through segments of society, by way of direct acquaintance, communication, media, and hearsay. Whether it ends up being performative will depend on how far it travels and on whether people believe it and deem it important enough to act upon it. For instance, the imagined prediction above would have been less performative, or not performative at all, if it hadn't been so widely disseminated, or if people hadn't come to believe in it or acted upon it by quickly selling their shares. It is thus evident that one and the same type of prediction can bring about performative effects of varying strength, or even entirely different effects. This suggests that performativity depends not only on the content of a representation issued but on something else, namely its *capacity* to influence the relevant outcomes, and this is what will be called its "performative power."[5] My conceptual proposal then, to repeat, is to think of any performative effect as a function of a representation and its corresponding performative power.

a) What makes for performative power? Whether and by how much a representation influences the course of events will depend on how many and who will become acquainted with it, and on whether and to what extent those reached will also believe and act upon it in ways that can change what the representation is about. For instance, Morgenstern, focusing on economic predictions of the business cycle, scrutinized different cases ranging from predictions that travel *throughout the economy* and enjoy *full authority*, all the way to predictions that only reach parts of the economy and are only partially believed, and he clearly saw that the strength of the ensuing effects, and thus, the performative power of the prediction in question, will substantially differ from case to case (see 1928, 99-102). Morgenstern also considered that there may be *competition* between predictions: various, mutually inconsistent predictions pulling the ensuing effects in "different directions", where the final outcome will depend on the outcome of this competitive process.[6]

Even though I follow Morgenstern's analysis, I take it that there are factors other than reach and acceptance that will also influence the performative power of a representation, in particular, its *relevance* or *importance* to people (e.g. are my savings at stake, as in the example above? Is my health affected? Etc.). For instance, consider a prediction that a particular cryptocurrency is soon going to shoot through the roof, but suppose that, for whatever reason, the prediction primarily reaches people that are very unlikely to invest in cryptocurrencies (e.g. age groups that tend to be uninterested in cryptocurrencies). So, even though the prediction travels through some segments of society, and it may be fully believed by those reached, it

[5] I borrow this term from Hardt et al. (2022) who use performativity to reason about the power of machine learning systems. Their use of "performative power" has a technical meaning (roughly, this refers to the largest change to a population that an algorithm can cause), which is closely related to my use of the term.

[6] Such cases may give rise to interesting power relations, for instance, when the issuing of one prediction weakens the performative power of another prediction. Such rival predictions will become relevant in the next section.

doesn't relate to their incentives and behavior in relevant ways so that the prediction ends up being not performative at all. What this shows is that relevance is necessary for a representation to enjoy high performative power, just like reach and acceptance are.

Taking these aspects together, performative power can be characterized as the *capacity of a representation to influence what it aims to represent, where this capacity is actualized by a tripod of reach, acceptance, and relevance*. In any real-world case, this tripod will be constituted by a complex concatenation of events and combination of factors: how a representation is issued and displayed, and the authority of the scientists issuing it; whether the representation is credible and whether people deem it important enough to act upon it; the way in which it is depicted in the news and shared on social media; whether *other* people seem to believe and act on it; how it will be passed on between subjects, etc. This leads to the next question to consider:

b) What carries performative power and how is it generated? When conceptualizing performativity as a function of a representation and its performative power, there is an important type-token distinction to be made. On my account, a scientific representation is any *type* of prediction, model, theory etc. that will be issued and disseminated through tokens of that same prediction, model, or theory. Examples of tokens are: someone reading a published paper containing a prediction on her device; someone else opening the same prediction on his device; a third person hearing about the prediction from a friend, or seeing it on social media – these are all tokens of one and the same type of prediction. It is apparent, then, that it is not a *token* representation (e.g. the prediction on my device) being performative; rather, performativity is constituted by the totality of reactions to a type representation.[7] It is thus also a type of representation that enjoys performative power, but its power will depend on and "trickle down" through tokens of that type of representation (or through tokens of similar types of representations, as will be discussed in the following section). Since performative power should be attributed to types of representations, for the sake of brevity, I will in the following use "representation" (and "prediction", "model", etc.) to refer to types, not tokens. When a (type of) representation has high performative power, tokens will tend to be passed on and generated anew, such as, for instance, when a person reads a prediction on her device, finds it relevant and credible, and then goes on to tell friends or family about it, thus generating a novel token of it. The performative power of a (type of) representation will tend to increase as more tokens are generated and passed on, in a process that boosts its reach and, in many cases, its acceptance and relevance.

The metaphorical depiction in Figure 1 might be found intuitive: think of the wine as a type of representation, and each filled glass as a token of that same representation. In the metaphor, glasses spill over and pass on the representation to other glasses in the next line below, thus generating novel tokens: the type trickles down through its tokens.

[7] I take it that this is the key difference to Austinian performativity, where it is a token utterance ("I now pronounce you husband and wife", "I bet…", "I promise…") that brings into being what it pronounces.

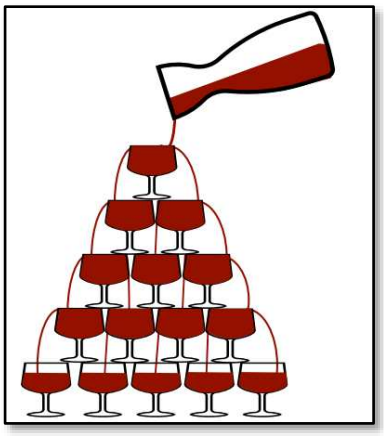

Figure 1: Metaphorical representation: a type "trickling down" through its tokens.

c) Finally, *how can performative power be measured?* Since this is the ability of a representation to influence the course of events, we can ask "how far" the issuing of the representation pushes outcomes away from those that would otherwise materialize. Accordingly, a measure of performative power will estimate the deviation from the counterfactual case in which the representation in question wouldn't be issued: the greater the deviation, the greater the performative power of the representation. As an example, consider a performative prediction of the likelihood of a recession in a particular country. This is shown on the left-hand side of Figure 2, where the X-axis represents the predicted value ranging from a 0 percent chance of recession to recession with certainty; and the Y-axis shows the actual likelihood of a recession occurring. The dashed line represents the baseline, that is, the likelihood of a recession if no prediction was issued. Each of the colored graphs represents a particular predictor. The blue predictor is performative in a slightly self-fulfilling way – for low values of the prediction, the likelihood of a recession occurring is a bit lower than baseline, and for high values, the likelihood is higher than baseline. The gridded area shows its total deviation from baseline. The red predictor is, likewise, self-fulfilling, albeit in a more extreme way. The combined lined and gridded areas show its total deviation from baseline; since this is greater than the gridded area alone, the red predictor has a higher performative power than the blue one.

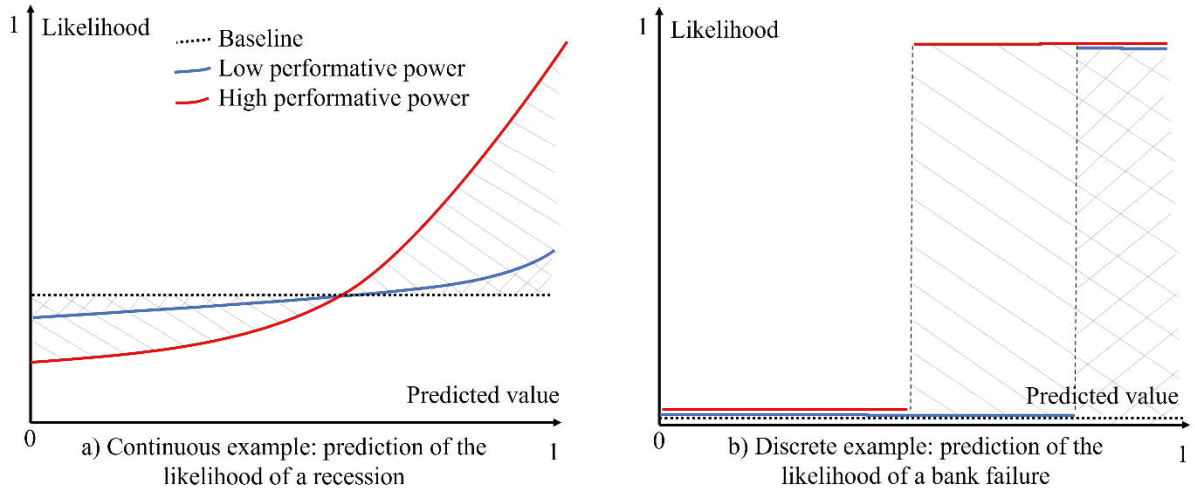

Figure 2: Performative prediction. Left-hand side: continuous example; right hand side: discrete case.

In other cases, performativity isn't generated continuously, but rather, there can be tipping points, the passing of which generates a critical mass of reactions triggering an "avalanche", a sudden strong effect. The prediction of a bank failure that becomes self-fulfilling is a case in point, as banks have sufficient liquid capital to pay out a number of customers, but when a critical mass of customers demand their money all at once, a bank will then fail to pay them out and will thereby go bankrupt. Such a case is shown on the right-hand side of Figure 2. The dashed line, again, shows the baseline, which is a very low probability of the bank failing. For the red predictor, there is a tipping point at around one half, which, if exceeded, triggers a run on the bank, which makes its failure almost certain. Similarly for the blue graph, for which, however, the threshold for triggering this event is higher. Since reactions are more sensitive to the red predictor, it has higher performative power than the blue predictor, as measured, as before, by the greater size of the combined lined and gridded areas. While the difference between continuous and discontinuous cases matters epistemically (Grunberg and Modigliani 1954; Simon 1954; this will be discussed in the next section), note, for now, that both in continuous and in discontinuous cases, performative power can be measured by estimating the deviation from baseline; the greater the deviation, the higher the power of a representation.[8]

To briefly sum up, performativity can be conceptualized as a function of a representation and its corresponding performative power, where the latter depends on how tokens of that representation trickle down through segments of society, thus determining its reach, acceptance and relevance, and it can be measured by estimating how far it "pushes" outcomes away from the baseline. But besides being a useful tool for conceptualizing performativity, performative power also has important ramifications for the evaluation of performative science: while there is nothing inherently problematic with performative power, as with economic or political power, we should be concerned when it becomes overly concentrated, or monopolistic. As I will argue next, concentrated power should also take center stage when thinking about performative science.

4. Concentrated Power

With the concept of performative power at hand, we can now put it to use for the evaluation of performative science. As I will argue in this section, there is no absolute value at which performative power becomes problematic, as even high power can be justifiable, namely *when the epistemic quality of the representation in question is equally high*. But this isn't always the case – then we have *concentrated power*. Power concentrations are problematic for epistemic and ethical reasons and they should be broken, but doing so can bring about different and, in some cases, more performative effects, and it follows that we shouldn't regard performativity as such as the problem, but rather, "junk" performativity resulting from power concentrations.

[8] Measuring performative power will in practice be difficult, in particular because this involves evaluating counterfactuals, for instance, estimating the baseline in which no prediction was issued in a scenario in which a prediction was actually issued and turned out to be performative. This might be possible by means of statistical and econometric methods, such as synthetic controls. Moreover, performative power has recently been operationalized in different settings in the field of machine learning (see Hardt et al. 2022; Hardt and Mendler-Dünner 2023), and we may hope that this can be adapted to the case of performative power in science. I leave it for future work to make progress on measurement; the difficulty in measuring performative power, in any case, doesn't affect the point of this paper, which is to introduce a useful conceptual framework for assessing performative science. I thank an anonymous reviewer for pressing me on this point.

There are two ways in which performative power can be overly concentrated: first, what I call a "monopoly" is an individual representation that enjoys high performative power while being epistemically deficient; second, a "tyranny of the majority" is a number of uniform representations that jointly (but perhaps not individually) have high performative power while being epistemically deficient. A monopoly can arise, for instance, when a scientist, or a small group of scientists, come to have a disproportionally "loud voice" on certain issues, which may be due, for instance, to their institutional affiliations, charisma, or the right contacts among policy makers and the media. Their "loud voice" implies that tokens of their representations will be effectively generated and passed on and, in many cases, believed and acted upon, thus yielding these representations a large amount of performative power, which, however, doesn't align with their epistemic credentials. By contrast, where large groups of scientists issue uniform kinds of representations, this may similarly generate high performative power, even when, individually, each representation wouldn't be overly performatively powerful, and when these representations are epistemically wanting, the result is a tyranny of the majority. Note that this doesn't require identity of the representations in question; similarity will often suffice, as similar kinds of representations will tend to reinforce a given effect, as in the real-world example below. Importantly, both monopolies and tyrannies will tend to marginalize the voices of other scientists and the power of their competing representations, even when these representations have better epistemic credentials.

While the distinction between monopolies and tyrannies is conceptually useful, in many real-world cases, they go hand in hand. Consider the now infamous case of Harvard economists Carmen Reinhart and Kenneth Rogoff, who purported to show that public debt above 90% of GDP is consistently associated with negative economic growth (2010). Their findings received a lot of attention (they trickled down through significant segments of society), and they were used to justify fiscal austerity policies around the world in the aftermath of the global financial crisis (cf. Rodrik 2015, 76) – before then-graduate student Thomas Herndon and co-authors found a spreadsheet error and various methodological flaws in the original paper, which severely challenged their results (Herndon et al. 2013). If Reinhart and Rogoff's results were among the factors that led to austerity policies being implemented, their case clearly qualifies as a monopoly – a representation that is performatively powerful while epistemically wanting. But, before Herndon et al. shook the evidence, Reinhart and Rogoff's findings were also validated by numerous other economists in a range of contexts (e.g. Kumar and Woo 2010; Cecchetti et al. 2011; Baum et al. 2012; Checherita and Rother 2012; Padoan et al. 2012), which, plausibly, increased their reach and reinforced their impact on policy and the public imagination. If this reading is correct, the case also qualifies as a special case of a tyranny, where a flawed result became individually powerful but was also followed by various concordant results, which only reinforced its power.

What distinguishes legitimate from concentrated performative power is that representations here come to enjoy high levels of power but are at the same time *epistemically deficient*, that is, their dissemination generates inaccurate and biased beliefs. Implicit in this account is that performative representations *can be* of high or low epistemic quality. But this assumption needs to be defended; as we saw in Section 2, there are concerns that, when the very publication or dissemination of a representation affects its fit with its target, this severely impairs its epistemic integrity. If a performative representation is regarded as necessarily epistemically deficient, this would make redundant the notion of concentrated power. However, this assessment would be too strong: philosophers have increasingly investigated how the

representational capacity of science can be preserved while acknowledging the inevitability of performative effects in various domains (e.g. Jimenez-Buedo and Guala 2016; Jimenez-Buedo 2021; Runhardt 2021; Vergara-Fernández et al 2023),[9] and I follow their assumption that a representation isn't necessarily epistemically doomed when being performative. While there are various plausible ways of evaluating the epistemic import of performative representations (van Basshuysen 2025 provides an overview), for present purposes, Grunberg and Modigliani's approach provides an illuminating example (1954; see also Simon 1954). Focusing on performative predictions of the business cycle, Grunberg and Modigliani asked under what conditions it is possible to align a prediction with the behavioral responses that will be triggered by its dissemination. For instance, in Figure 2 in the previous section, their approach would amount to searching for equilibria in which the predicted value and the likelihood of the respective outcome (i.e. the X- and the Y-values) are identical, and their main result is that such an equilibrium always exists in the continuous case (i.e. on the left-hand graph but not necessarily on the right-hand graph). Inspired by Grunberg and Modigliani's approach, recent work in machine learning has extended and operationalized their result, by providing algorithms for finding predictions that actively shape social outcomes while at the same time accurately predicting these outcomes (see Perdomo et al. 2020; Hardt and Mendler-Dünner 2023; Perdomo 2025).

The rationale for the approach inspired by Grunberg and Modigliani is that a prediction that aligns predicted with actual outcomes would be better, epistemically speaking, than one that doesn't, other things being equal. Since it is in many cases possible to find such a prediction, it follows that there is such a thing as the epistemic import or quality of performative representations: even when representations do alter their targets, it doesn't follow that anything goes, but rather, we should (and often can) aim to align them with their targets.[10] But this approach also has implications for the evaluation of performative science: since the equilibria aligning predicted with actual outcomes will typically *differ* from baseline, this means that, when following this approach, we aim not to mitigate performative effects, but rather, to control or steer them in particular ways (see Jimenez-Buedo and Guala 2016 for a related argument). Since predictions that are better, epistemically speaking, do generate performative effects, this shows that the epistemic problem shouldn't be attributed to performativity as such, but rather, to junk performativity generated by power concentrations.

Power concentrations are not only epistemically problematic, but acting on the inaccurate and biased beliefs generated by them can also have ethically and socially undesirable consequences. This is the grain of truth in Winsberg and Harvard's argument against treating

[9] I thank an anonymous reviewer for pointing me to this literature.

[10] This approach faces two challenges. First, as we have seen, in discontinuous cases, an equilibrium might not exist, and even when it does, Grunberg and Modigliani's result doesn't guarantee that we'll be able to find it. This is a technical challenge that we may hope to tackle, for instance, by means of approximate equilibria, and better algorithms for finding equilibria. This line of work is currently being conducted in the field machine learning, where the original result of Grunberg and Modigliani has been generalized and it has been shown that the continuity requirement of the reaction function can be dropped (see Perdomo 2025). This makes the task of finding accurate performative predictions more practicable. Second, it might be argued that aligning actual with predicted outcomes doesn't really "solve" the epistemic problem posed by performativity because the prediction might then be right for the wrong reason, that is, descriptively good only "because it *shaped* the target well" (cf. Marchionni 2024, 18, emphasis in original; see also [manuscript blinded]). This suggests that there might be *more stringent epistemic criteria* than Grunberg and Modigliani's (which I explore more fully in van Basshuysen 2025), which, however, confirms my point that performative representations *can be* more or less epistemically good.

performativity as a model purpose (2022), which we encountered above: they are correct in assuming that performative effects *can* sometimes impair the accurate weighing of evidence and thereby preclude an impartial cost-benefit analysis. But the problem shouldn't have been attributed to performativity across the board (see van Basshuysen 2023), but rather, to junk performativity resulting from monopolies and tyrannies that feed into biased beliefs and decisions. For instance, uniform predictions downplaying the risks associated with certain kinds of behaviors might end up generating beliefs that feed into status quo bias, that is, the propensity to stick to these behaviors, even when the situation would require a strong reaction. Breaking power concentrations in such cases might alter the ensuing performative effects, by bringing about stronger effects, which can be socially quite desirable. In contrast to the epistemic aspects discussed above, it would be a mistake, however, to think that concentrated power will by necessity kick off undesirable performative effects. Imagine, for instance, an influential but overblown warning of some imminent threat, which triggers behavioral changes impeding the threat from materializing, whereas a more realistic description of the situation would have failed to trigger sufficient behavioral changes. Sometimes overblown statements may have better consequences, in a social or ethical sense, than more accurate ones because behavioral responses can be sluggish, that is, individuals may fail to react to information in ways that would be adequate to the situation in question. Nevertheless, we may hope that, on average, breaking monopolies and tyrannies will produce more desirable effects because this offers itself to rational decision making leading to desirable outcomes, which would otherwise be mere flukes.

To briefly sum up this section, we shouldn't regard performativity as such as the problem but, rather, that representations come to enjoy monopolistic or tyrannous power over performative effects, and we should worry about biased beliefs and decision making resulting from such power concentrations. Our strategy should accordingly be to get rid of power concentrations, but not necessarily to get rid of performative effects; on the contrary, in some cases, eradicating power concentrations may involve eliciting different, or even more, performative effects. Performative representations have a productive role to play in a democratic society if we learn how to deal with problematic power concentrations and resulting biases. This is the topic to which we turn next.

5. Managing Performativity in a Democratic Society

Performative power that is concentrated in the hands of individual scientists, or scientific disciplines, is much more than a conceptual possibility but is, rather, a very real concern following from science's increasingly influential role in society. For instance, Diane Coyle, arguing that ideas and conceptualizations from economics and economists' influence on society and policy making significantly shape the social world, worries about harmful societal effects ensuing from concentrated influence and a lack of diversity in economics (2021). And in the context of the COVID-19 pandemic, Lohse and Canali argue that epidemiology as a discipline was too dominant in informing policies and that perspectives from the social sciences should be heard and acted upon more to reach balanced policy making in public health emergencies (2021). But how might such undesirable power concentrations be broken, or be prevented before they even constitute? This is the question of how performative science ought to be managed in a democratic society – which, I believe, is a big project that is perhaps comparable in terms of the attention it requires to that of values in science, to which philosophers have dedicated considerable attention in the last decades (e.g. Longino 2002; Douglas 2009; Elliott 2017; Holman and Wilholt 2022). In contrast, the normative evaluation and management of

performativity are only beginning to be discussed in the literature (e.g. Godman and Marchionni 2022; Khosrowi 2023; van Basshuysen 2023, 2025; Ortmann 2025). Here, I will only grapple with an important precondition for making progress on the bigger project of evaluating and managing performative science. This precondition concerns the question of how responsibility for performative effects should be conceptualized and allocated in a democratic society.[11]

If we view any performative effect as a function of a representation issued and its corresponding performative power, which is the conceptual framework I have argued for here, we face a twofold task, consisting of, first, allocating responsibility for representations, and, second, allocating responsibility for performative power. My suggestion, then, is a broad division of labor: individual or groups of scientists have responsibility for their representations, including the anticipation of any detrimental performative effects these may generate (cf. Godman and Marchionni 2022); whereas responsibility for performative power should be distributed more broadly in a democratic society. Let's consider these two parts in turn.

First, while it seems obvious that scientists hold responsibility over the predictions they issue, importantly, this responsibility covers not only the epistemic import of their predictions (as discussed above), but it should also include any detrimental performative effects potentially accruing from them. As Godman and Marchionni argue, scientists should anticipate these effects and, where possible, mitigate or even preclude them from materializing in the first place (2022).[12] They discuss the case of research into gender differences in cognitive abilities, which threatens to reinforce undesirable gender stereotypes. When this happens, they suggest that scientists should then reform communication strategies to avoid such detrimental effects from occurring (2022, 11). Similarly, Ortmann and Veit consider, in the context of climate change mitigation, a modeling choice where one way of describing the problem situation will result in a Tragedy of the Commons, whereas the other will result in a cooperative solution (2023). If both modeling choices are equally self-fulfilling, the preferred choice, according to them, is clearly to generate the cooperative outcome. Similarly, various cases have been discussed in which a specific evocation of performative effects might be seen as socially desirable, for instance, when they help strengthen indigenous cultures and lessen harmful stereotypes, as argued by Koskinen (2022), or when they contribute to increased social distancing behavior in a pandemic (see van Basshuysen et al. 2021). In addition, as Ortmann (2025) argues, it is often impossible to escape making performative choices on behalf of others' well-being, and whenever this is the case, the likely performative effects accruing from the issuing of a given representation will have to be considered.

In many cases, however, there may be disagreement about what kinds of effects should be thought of as desirable or undesirable, so that, what the preferable choice of action is will not be so obvious. In typical cases, there will be winners and losers from the evocation of given effects, which presents us with the question of what legitimizes the evocation of some effects,

[11] My recommendations in this section fall short of providing advice that is directly actionable. Rather, my aim is to raise the issue of responsibility for performativity, and to provide a preliminary discussion. I thank an anonymous reviewer for raising this point.

[12] Godman and Marchionni discuss and reject various reasons against this view. A possible reason to think scientists might be excused for not taking into account performative effects accruing from their research choices is that the respective effects are so unpredictable that scientists cannot meaningfully be held responsible for them. See e.g. Hacking (1995; 1999) for this line of argument. However, against Hacking, Godman and Marchionni convincingly argue claims that performative effects cannot be meaningfully anticipated and managed are exaggerated (2022; see also Marchionni, Zahle and Godman 2024).

rather than others. There are various possible ways in which this question might be answered; taking clues from the literature on values in science, one answer would be to require scientists to base research decisions potentially bearing on performative effects on shared values and democratically endorsed aims (cf. Intemann 2015). Alternatively, we could think of scientists as representatives who are licensed to make certain consequential (including value-laden) decisions on behalf of the citizenry (Fernández-Pinto and Hicks 2019; Hilligardt 2023), and that this license should cover any performative effects accruing from their research. Irrespective of which approach is deemed more plausible, emphasis will be placed on inclusive democratic deliberation, while allowing scientists to act as stewards for their performative effects, aiming to foster socially desirable performativity while counteracting undesirable instances.

Second, how should responsibility for performative power be distributed? Remember that this depends on a tripod of reach, relevance, and acceptance. Since the tripod will be determined by a combination of the issuing scientists' position within their communities and society more broadly, as well as on societal and policy demand for models and predictions about specific issues, it will often be beyond scientists' ability to meaningfully anticipate and control for these factors. Think, for example, of the sudden demand for epidemiological forecasts in the early COVID-19 pandemic, which epidemiologists could hardly foresee or wish for. This suggests that individual scientists' responsibility for the performative power of their representations should be regarded as limited (even though not as nil, for see below). It would also be implausible to expect individual scientists to restrict their own performative power to prevent power concentrations, in particular where they are under pressure by departments and funding bodies to show the policy impact of their research, as many increasingly are. This provides additional reason against the view that scientists should be held solely responsible for their performative power. Accordingly, responsibility for removing power concentrations shouldn't depend on individual scientists alone; rather, this should be distributed more broadly, to include research institutions and departments at universities, professional networks, science journals, science activist groups, science journalism and the press, and political institutions. Each of these stakeholders should strive to provide an inclusive institutional framework within which competition for influence and the best ideas can take place in a way that prevents monopolies and tyrannies from occurring, or which at least makes their occurrence less likely.

Unfortunately, approaching this ideal immediately raises significant epistemic and action-related problems. The epistemic problem is that it will be difficult, in practice, to gauge when the performative power of a representation is overly concentrated. For instance, there may be differences in the influence of (individual or groups of) scientists, which may lead to their representations being more or less performatively powerful, which are, however, an expression of a proper epistemic competition between scientists. There is a difficulty to state precisely to what extent performative power is legitimate and at what point it becomes an illegitimate power concentration. While I cannot fully answer this question here, it is clear that each scientist being equally influential would itself be a distortion, and similarly, each view being equally represented would be a distortion (as it is sometimes reasonable that even a large proportion of scientists agree on certain views – that there is human-induced global warming is a case in point). Thus, strict equality won't do as a requirement for excluding concentrated power. We can, perhaps, find the beginning of a better requirement in Helen Longino's idea of "tempered equality". While this notion accepts that there can be differences in intellectual capacities, training, etc., it demands that contextual factors, in particular social position and economic power, ought not determine whose views are taken seriously within a community (2002; 131-

134). Similarly, I suggest that scientists' influence *outside* their communities, which will determine the performative power of their representations, should be based on merit alone and should not be decisively influenced by contextual factors, such as membership in a socially dominant group.

But even when we come to know precisely when there will be a monopoly or a tyranny, there remains an action-related problem, namely, that it will in many cases be hard to act on this knowledge to preclude these power concentrations from occurring, or to break them after the fact. Due to the complex nature of how token representations trickle down through segments of society, it would be difficult or impossible to track and control for the effects of any particular representation. Rather than focusing on individual representations becoming performative, it might, as an approximation, be better to focus on the institutional framework that might facilitate the occurrence of power concentrations, and on (individual or groups of) scientists as representatives for their representations issued. Accordingly, strategies to remove power concentrations and to achieve a level playing field might often consist in a reorganization of scientific communities, for instance, to closer resemble relevant constituencies. Concrete measures to achieve this may include increasing diversity in hiring decisions at departments, accepting diverse methodological approaches to be published in pertinent journals and taught in curricula, and diversifying the voices that get heard in science communication and political consultancy, amongst others – but, it's clear that this is only a preliminary list, and future work will have to detail the measures and institutional frameworks that might mitigate power concentrations. A final problem that can only be mentioned here is that the neat division of labor suggested here is significantly complicated, in practice, by the fact that representations and their power can be interrelated in intricate ways. In particular, a prediction, or specific ways in which it is communicated, can in turn influence its performative power. For instance, concerning the policy and public impact of academic economists, Coyle writes that "[t]he impact is delivered, the eye of the public is caught, by confident statements of extreme views, rather than by modest or nuanced analyses of complicated situations" (2021, 70). When predictions themselves, or ways of expressing or framing them, influence performative power in this way, scientists, being responsible for their predictions, will then also be in part responsible for their ensuing power, but it may, in practice, often be difficult to detect and to avert manipulations of performative power.

6. Conclusion

I have argued for a novel conceptualization of performativity, which naturally suggests a fundamental re-evaluation of performative science. Once we view performativity as a function of a representation and its concomitant performative power, it becomes clear that it's not performativity as such that is problematic (contrary to what is still often assumed); rather, we should worry that some representations come to enjoy great performative power while being epistemically wanting, and that this can generate biased beliefs and socially undesirable outcomes. Our efforts should thus concern breaking power concentrations, which may, in some cases, well result in more performativity, and it follows that there is a productive role for performativity if we learn how to demarcate legitimate from illegitimate ways of dealing with performative power. Finally, I have made some preliminary suggestions about how this might proceed, by focusing on the question of how responsibilities for performative power and predictions should be attributed in a democratic society. Yet this line of thought is in its infancy and the author would welcome an inclusive debate on this question.

References

Baum, Anja, Cristina Checherita-Westphal, and Philipp Rother. 2012. "Debt and Growth: New Evidence for the Euro Area." European Central Bank Working Paper Series No 1450

Buck, Roger. 1963. "Reflexive Predictions." *Philosophy of Science* 30(4):359-369.

Cecchetti, Stephen, M. Mohanty and Fabrizio Zampolli. 2011. "The real effects of debt." BIS Working Papers No. 352.

Checherita-Westphal, Cristina, and Philipp Rother. 2012. "The impact of high government debt on economic growth and its channels: An empirical investigation for the euro area." *European Economic Review* 56:1392-1405.

Coyle, Diane. 2021. *Cogs and Monsters: What Economics Is, and What It Should Be*. Princeton and Oxford: Princeton University Press.

Douglas, Heather. 2009. *Science, Policy, and the Value-Free Ideal*. Pittsburgh, USA: University of Pittsburgh Press.

Elliott, Kevin. 2017. *A Tapestry of Values: An Introduction into Values in Science*. Oxford: Oxford University Press.

Fernández-Pinto, Manuela, and Daniel J. Hicks. 2019. "Legitimizing Values in Regulatory Science." *Environmental Health Perspectives* 127(3). https://doi.org/10.1289/EHP3317

Godman, Marion, and Caterina Marchionni. 2022. What should scientists do about (harmful) interactive effects? *European Journal for Philosophy of Science* 12(63). https://doi.org/10.1007/s13194-022-00493-7

Grünbaum, Adolf. 1956. "Historical Determinism, Social Activism, and Predictions in the Social Sciences." *British Journal for Philosophy of Science* 7(27):236-240.

Grunberg, Emile, and Franco Modigliani. 1954. "The Predictability of Social Events." *Journal of*
Political Economy 62(6):465-478. https://doi.org/10.1086/257604

Guala, Francesco. 2007. "How to Do Things with Experimental Economics." In *Do Economists Make Markets? On the Performativity of Economics*, ed. Donald MacKenzie, Fabian Muniesa, and Lucia Siu, 128-62. Princeton, NJ: Princeton University Press

Hacking, Ian. 1995. "The looping effects of human kinds." In D. Sperber, D. Premack and A.J. Premack (Eds.), *Causal cognition: A multidisciplinary debate* (pp. 351-394). Oxford: Clarendon Press.

Hacking, Ian. 1999. *The Social Construction of What?* Cambridge, MA: Harvard University Press.

Hardt, Moritz, Meena Jagadeesan, and Celestine Mendler-Dünner. 2022. "Performative Power." *arXiv*:2203.17232

Hardt, Moritz, and Celestine Mendler-Dünner. 2023. "Performative Prediction: Past and Future." *arXiv*:2310.16608v1

Herndon, Thomas, Michael Ash, and Robert Pollin. 2013. "Does High Public Debt Consistently Stifle Economic Growth? A Critique of Reinhart and Rogoff." Amherst: University of Massachusetts at Amherst, Political Economy Research Institute, April 15.

Hilligardt, Hannah. 2023. "Partisan science and the democratic legitimacy ideal." *Synthese* 202:135.

Holman, Bennett, and Torsten Wilholt. 2022. "The new demarcation problem." *Studies in History and Philosophy of Science* 91:211-220.

Intemann, Kristen. 2015. "Distinguishing between legitimate and illegitimate values in climate modeling." *European Journal for Philosophy of Science* 5:217-232.

Jiménez-Buedo, María. 2021. "Reactivity in social scientific experiments: What is it and how is it different (and worse) than a Placebo effect?" *European Journal for Philosophy of Science* 11(2), 42.

Jiménez-Buedo, María, and Francesco Guala. 2016. "Artificiality, reactivity, and demand effects in experimental economics." *Philosophy of the Social Sciences* 46(1):3-23.

Koskinen, Inkeri. 2022. "Reactivity as a tool in emancipatory activist research." *European Journal for Philosophy of Science* 12:65. https://doi.org/10.1007/s13194-022-00487-5

Kumar, Manmohan S., and Jaejoon Woo. 2010. "Public debt and growth", IMF Working Paper, No. 10/174.

Lohse, Simon, and Stefano Canali. 2021. "Follow *the* science? On the marginal role of the social sciences in the COVID-19 pandemic." *European Journal for Philosophy of Science* 11:99. https://doi.org/10.1007/s13194-021-00416-y

Longino, Helen. 2002. *The Fate of Knowledge*. Princeton University Press.

MacKenzie, Donald. 2006. *An engine, not a camera. How financial models shape markets*. Cambridge: MIT Press.

Marchionni, Caterina, Julie Zahle, and Marion Godman. 2024. "Reactivity in the human sciences." *European Journal for Philosophy of Science* 14(8).

Merton, Robert K. 1948. "The Self-Fulfilling Prophecy." *The Antioch Review* 8(2):193-210.

Merton, Robert K. 1949. *Social Theory and Social Structure*. The Free Press.

Morgenstern, Oskar. 1928. *Wirtschaftsprognose: Eine Untersuchung ihrer Voraussetzungen und Möglichkeiten*. Berlin & Heidelberg: Springer. https://doi.org/10.1007/978-3-7091-2157-3

Ortmann, Jakob. 2025. "Performative paternalism." *European Journal for Philosophy of Science*, 15:25.

Ortmann, Jakob, and Walter Veit. 2023. "Theory roulette: choosing that climate change is not a tragedy of the commons." *Environmental Values* 32(1): 65-89. https://doi.org/10.3197/096327122X16452897197784

Padoan, Pier Carlo, Urban Sila, and Paul van den Noord. 2012. "Avoiding debt traps: financial backstops and structural reforms." OECD Economics Department Working Paper No. 976

Perdomo, Juan Carlos, Zrnic, Tijana, Mendler-Dünner, Celestine, and Moritz Hardt. 2020. "Performative Prediction." *Proceedings of the 37th International Conference on Machine Learning*, PMLR 119:7599-7609

Perdomo, Juan Carlos. 2025. "Revisiting the Predictability of Performative, Social Events." *2025 International Conference on Machine Learning (ICML)*

Popper, Karl. 1957. *The Poverty of Historicism*. Boston: Beacon Press.

Reinhart, Carmen M., and Kenneth Rogoff. 2010. "Growth in a time of debt." *American Economic Review* 100(2):573-78.

Rodrik, Dani. 2015. *Economics Rules*. Oxford, UK: Oxford University Press.

Runhardt, Rosa W. 2021. Reactivity in measuring depression. *European Journal for Philosophy of Science* 11(3), 77.

Simon, Herbert A. 1954. Bandwagon and Underdog Effects and the Possibility of Election Predictions.
Public Opinion Quarterly, 18(3), 245. https://doi.org/10.1086/266513

van Basshuysen, Philippe. 2023. "Austinian Model Evaluation." *Philosophy of Science*, 90(5):1459-1468, 2023. https://doi.org/10.1017/psa.2023.24

van Basshuysen, Philippe. 2025. "Performativity in Science: Past and Future." *Philosophy Compass* 20:e70062. https://doi.org/10.1111/phc3.70062

van Basshuysen, Philippe, Lucie White, Donal Khosrowi, and Mathias Frisch. 2021. "Three Ways in Which Pandemic Models May Perform a Pandemic." *Erasmus Journal for Philosophy and Economics* 14(1): 110-27. https://doi.org/10.23941/ejpe.v14i1.582

Vergara-Fernández, Melissa, Conrad Heilmann, and Marta Szymanowska. 2023. "Contextualist model evaluation: models in financial economics and index funds." *European Journal for Philosophy of Science* 13:6. https://doi.org/10.1007/s13194-022-00506-5

Winsberg, Eric, and Stephanie Harvard. 2022. "Purposes and duties in scientific modelling." *Journal of Epidemiology and Community Health* 76:512-517.