

Appraising the Epistemic Performance of Social Systems: The Case of Think Tank Evaluations

by François Claveau and Andréanne Veillette

Department of Philosophy and Applied Ethics, Université de Sherbrooke

Abstract

This article elaborates a conceptual framework to systematize the epistemic evaluation of social systems. This framework can be used to structure an evaluation or to characterize and assess existing ones. The article then uses the framework to assess four representative evaluations of think tanks. This meta-evaluation exemplifies how the framework can play its structuring role. It also leads us to general conclusions about the existing evaluations of think tanks. Most importantly, by focusing on the organizational level, existing evaluations miss factors that are situated at the network and at the ecosystemic levels and that significantly determine how well think tanks serve society in producing and disseminating knowledge relevant to public policy. This conclusion suggests the need for epistemic evaluations of think tank ecosystems.

Keywords. systems-oriented social epistemology; think tanks; applied epistemology; epistemic evaluations

1 Introduction

Social systems can be studied and evaluated¹ from an epistemic perspective. Many contemporary epistemologists are busy doing exactly that. They evaluate, for instance, the legal system (Laudan 2006), Wikipedia (Fallis 2008), the education system (Kotzee 2013), medicine (Bluhm and Borgerson 2019) and, quite extensively, academic science (Longino 1990). Yet, much work remains to be done to systematize the epistemological study of social systems.

A decade ago, Alvin Goldman called this study the “least familiar and most adventurous form of [social epistemology].” (Goldman 2010, 190) All fields of knowledge start adventurous, but they mature when their practitioners converge on frameworks of inquiry, frameworks that structure research projects and that can retrospectively be used to evaluate the strengths and weaknesses of existing work. However, the recent wave of epistemological studies of social systems has only proposed pieces of a framework, focusing on the question of what we should mean by the underlying notion of ‘epistemic performance.’² Strikingly, only passing remarks are made to the fact that the epistemology of social systems has a strong *empirical* component and should thus have guidelines to structure its handling of

1 Because this article is concerned with two levels of appraisal, we adopt the following convention in order to make it easier for the reader to follow: “assessment” will be reserved for higher level appraisals while “evaluation” will be reserved for lower level appraisals.

2 We put in this category, for example, Longino’s procedural objectivity, Goldman’s V-value, Fricker’s epistemic justice and Bishop and Trout’s strategic reliabilism. See section 3.

Article forthcoming in *Episteme*. Corresponding author: <Francois.Claveau@USherbrooke.ca>

Acknowledgements. The authors thank participants at the ICPP panel "Think Tank Discourse and Political Change" (Montreal, June 2019) and at the research retreat of the Canada Research Chair in Applied Epistemology (July 2019) for thoughtful comments on an early draft of this article. They acknowledge the financial support of Canada Research Chairs [950-230644] and of the Social Sciences and Humanities Research Council of Canada [435-2019-0863]. The authors have contributed equally to this article.

evidence and empirical generalizations.³ The first goal of this article is to contribute a complete framework which can be used for the epistemic evaluation of different social systems. We expect our framework to be criticized and improved upon as the field matures.

Although still somewhat adventurous, systems-oriented social epistemology represents an important area of study because its objects are, most notably, organizations which produce and disseminate knowledge (or knowledge claims) with society-wide impacts. These organizations play crucial roles in the day to day life of citizens, making them prime objects of epistemic evaluation. One such system which is in need of epistemological scrutiny is the think tank. Think tanks permeate the media where they often replace more traditional forms of expertise (Drezner, 2017) and produce research with the express goal of influencing public policy. Yet, to our knowledge, no work in social epistemology evaluates think tanks. A second major contribution of our article is to launch the social epistemology of think tanks.

What are think tanks? It is a truism to say that think tanks are hard to define: “The boundary line between these organizations and others is not clear-cut.” (Weaver 1989, 563) For our purposes, a think tank is an independent, nonprofit organization whose main function is to produce and disseminate public-policy studies and analysis. Under our working definition of a think tank, the requirements which must be fulfilled to qualify as independent are fairly minimal: the organization must be a separate legal entity, unaffiliated with the state, political parties, universities or lobby groups.

In fact, all think tanks depend on resources from different actors to thrive. In this sense, they are free from legally binding institutional ties, but they must maintain several informal ties to these very same institutions in order to prosper (Medvetz 2012, chap. 1). The varying types and strengths of these ties to other institutions result in significant heterogeneity among think tanks. To capture this variety, typologies of think tanks have proliferated. Most typologies recognize three moments in time which correspond to the three main types of think tanks: universities without students, contract research organizations and the advocacy tanks (Weaver 1989; Abelson 2000; Stone 2013). Other typologies make finer distinctions, recognizing up to nine different types of think tanks (McGann 2007; for a recent discussion of typological efforts see Landry 2018). For the purposes of this article, noting the diversity among think tanks is important, but agreeing on a typology is not.

One thing is consensual in the literature: “think tanks present themselves, and are represented by the media, as scientific establishments, composed of experts and scholars engaged in the task of thinking, writing and publishing” (Stone 2007, 261). Because think tanks explicitly claim to produce public-policy knowledge and because many think tank experts regularly take the place of academic experts in public discourse, their epistemic contribution to society must be investigated.

3 The history and philosophy of science (HPS) is a noteworthy exception to this claim, with a sustained discussion of the methodology of HPS that reaches back at least to Lakatos (1971) and up to today (e.g., Kinzel 2015; Scholl and Rätz 2016; McAllister 2018; Schickore 2018). This literature differs from the current project in being focused on science as a historical phenomenon, while we focus on contemporary social systems. Social epistemology has methodological lessons to draw from HPS, although the task of reinterpreting this methodological literature in light of the goals of social epistemology is best left for a separate article.

Many evaluations of think tanks already exist. However, their focus is not explicitly epistemic, perhaps because epistemologists have not yet taken up the challenge. We contend that it is necessary to approach the existing literature on think tank evaluations in a systematic way in order to find the areas where further efforts are needed. In what follows, we assess existing evaluations in order to see if they adequately evaluate the epistemic contributions of think tanks to society. To do this, we first elaborate a conceptual framework to systematize the epistemic evaluation of social systems – think tanks being only one example among many types of social systems. This framework is comprised of four necessary components: the level at which the evaluation takes place, the chosen conception of epistemic performance, the empirical adequacy and the practical relevance of the evaluation. We then apply the framework to a representative sample of existing evaluations of think tanks. Our assessment of evaluations – our ‘meta-evaluation’ if you will – indicates that the existing evaluations share a blindspot, i.e., their choice to evaluate the properties of think tanks instead of properties of their network and ecosystem. This conclusion indicates a promising direction for the nascent social epistemology of think tanks.

2 Different levels of socioepistemic systems

A system is a whole constituted of parts in interaction. Individual humans can thus be understood as systems and, since they have epistemic properties, they can be studied and evaluated as epistemic systems. But this article is about systems with epistemic properties at higher levels than that of individual humans, i.e., socioepistemic systems. Without claiming to exhaust types,⁴ we distinguish between three levels: organizations, networks of similar organizations and ecosystems.

Organizations are a specific type of social systems “that involve (a) criteria to establish their boundaries and to distinguish their members from nonmembers, (b) principles of sovereignty concerning who is in charge, and (c) chains of command delineating responsibilities within the organization.” (Hodgson 2006, 8) Typical examples of organizations include firms, political parties and universities. These systems are *formal* organizations in the sense that they have a legal identity, but any organization has a structure that gives it some degree of permanence. For instance, individual humans filling certain positions (e.g., the president or the treasurer) can change while the organization persists. An organization also has a sort of agency. On that basis, we can attribute purposes and commitments to an organization. Focusing on epistemic properties in particular, an organization can be said to be committed to certain claims and arguments.

Organizations come in types – firms, sport teams, research centers, etc. What we call a ‘network of similar organizations’ – ‘network’ for short – is a system composed of interacting organizations of the same type. For instance, think tanks interacting with other think tanks would make up a network of similar organizations whereas think tanks interacting with universities would not. Such a network is not

4 In social epistemology, much work is done on “groups” as doxastic agents – i.e., collective agents having belief-like attitudes (e.g., Wray 2001, Gilbert 2014). Groups can be understood as a special type of systems given our definition, although Alvin Goldman (2010) separates group epistemology from systems-oriented social epistemology. Furthermore, what we define as an organization is a type of group (doxastic) agent.

simply a higher level organization. For instance, an industry is a network of firms in the same economic sector that does not have the structural properties of an organization. It is possible that an organization represents and partly regulates a network – for instance, a league for a network of sport teams – but this organization is not identical to the network.

Finally, we draw a distinction between a network of similar organizations and an ecosystem.⁵ The discriminating factor is that an ecosystem is a system composed of a more diverse set of units than what we call a network. The analogy with a biological ecosystem works as follows: different types of socioepistemic units which interact in a given environment compose a socioepistemic ecosystem in the same way that different biological species interact in a given environment to compose a biological ecosystem. A think tank ecosystem thus includes, beyond think tanks, other types of organizations, such as the media, academia, political parties, bureaucracies and funders, all interacting in a given environment.

3 Evaluating epistemic performance

An epistemic evaluation is a specific type of system appraisal: the standard used in the evaluation is a conception of epistemic performance. Some evaluation protocols are better than others. In this section, we articulate considerations to take into account when assessing epistemic evaluations. The first subsection is about the selection of a conception of epistemic performance, an important topic in the existing literature. The second subsection focuses on issues with the empirical basis of the evaluation. The last subsection turns to concerns about the relevance of the evaluation. These last two subsections cover considerations little discussed in social epistemology to date.

3.1 Conceptions of Epistemic Performance

An epistemic evaluation is a type of appraisal where the object is valued according to knowledge-related conditions. The common denominator of all epistemic evaluations is that, instead of prioritizing, say, aesthetic or moral values, the values of attaining truths and avoiding errors take precedence.

Some epistemologists prioritize epistemic values to the point of almost excluding any other value. They propose *purist conceptions* of epistemic performance. Alvin Goldman's original formulation of a veritistic social epistemology is a case in point. In *Knowledge in a Social World* (Goldman 1999), Goldman builds a conceptual framework to evaluate the epistemic value of specific practices in a wide-range of domains such as science, democracy and education. An objective of the framework is to quantify to what degree some epistemic practices generate true beliefs and prevent the creation of false beliefs. Such a framework allows the comparison of the epistemic merit of different organizational choices.

5 Some prefer the term “ecology” to ecosystem (Lindquist 2003). Yet, the latter is etymologically closer to what we want to refer to: not the *logos* on a home, but the actual *systema* of it. We conjecture that ‘system’ has gained a bad flavor for many social scientists through its extensive use by functionalism. We think it is better to fight the semantic association between ‘social system’ and the idea that this entity functions well than to use words that are etymologically at odds with what we want to refer to.

A concern arises with this purist conception of epistemic performance: should a true belief be given the same weight regardless of its relevance? For instance, should the fact that Pauline believes correctly that ‘the colour of the tabletop is darker than the colour of the floor’ contribute in the same way in establishing the level of epistemic performance of her vision than her correct belief that ‘her head is directly in the trajectory of a fastball’? Undoubtedly, the stakes are higher when it comes to true belief in the second proposition because believing it can inform the decision to dodge and thus can make Pauline avoid a serious headache (or something worse). It is also assumed that Pauline’s interest in the relative brightness of surfaces is rather mild, perhaps she simply wanted to come up with a weird example in a paper she is writing. While the interest in believing both propositions is markedly different, true belief in both propositions would be weighed in the same way according to a purist conception of epistemic performance. This seems to be a problem for the purist conception.

Goldman initially replied to this concern by allowing for what he called a “moderate role” of interest, which he later recognized was closer to a “*minimal role*” (Goldman 1999, 95; 2000, 321). According to Goldman, the magnitude of interest in a question does not matter. The only constraint resting on the evaluated belief is that it must be an answer to “a question of interest” (Goldman 2000, 321). Goldman then nuanced his position in replies to commentators – for instance, by welcoming both “*pure veritistic epistemology* and *extended veritistic epistemology*” (Goldman 2002, 218). Accordingly an extended epistemology would study the veritistic properties of practices, but would rank practices on a more inclusive set of conditions. Since Goldman does not say much more on this non-purist alternative, we have to turn to work done by other epistemologists who further developed the extended conception of epistemic performance.

An example of such an extended conception is Bishop and Trout’s “mongrel epistemology” (Bishop and Trout 2017, 111).⁶ Their framework – dubbed *strategic reliabilism* – relies on three conditions for epistemic performance: robust reliability, efficiency and significance (Bishop and Trout 2005, 55). Robust reliability is understood as processes (or rules) which consistently give a high ratio of true judgments to total judgments over a large scope of environmental variations. Efficiency refers to the sparing of resources in successfully accomplishing tasks. Significance expresses the degree to which a question is worth spending resources on. The extended conception of epistemic performance at play here is that “epistemically excellent reasoning is efficient reasoning that leads in a robustly reliable fashion to significant, true beliefs.” (Bishop and Trout 2008, 1061)

How should epistemologists decide between a purist and an extended conception, and how should they further specify epistemic performance beyond this dichotomy? A full answer to this question will need to wait for another article, but our discussion in this subsection models how we think a decision can be reached: it is possible to have an exchange of arguments over what is deemed reasonable to include or exclude in the conception of epistemic performance for a given system. For instance, it is not quite reasonable to exclude “significance” when we aim to evaluate Pauline’s visual system. Furthermore, as we illustrate below in our discussion of think tanks, it seems to us that an appropriate conception of

6 See also (Fallis 2006) for a proposal to extend Goldman’s framework to more thoroughly take non-epistemic interests into account.

epistemic performance will be sensitive to context: an appropriate conception for one type of system will not necessarily fit for another type. Finally, we have no objection to a piecemeal and iterative approach to epistemic performance. Instead of spending years arguing over complete conceptions of epistemic performance, it is better to focus on some aspects that are arguably central to epistemic performance and use those to perform evaluations. The provisional results can later be improved by new rounds of conceptual and empirical work.

3.2 Empirical adequacy of the evaluation

An epistemic evaluation relies on empirical research to determine the extent to which the system meets the chosen conception of epistemic performance. There are various factors threatening the success of the empirical part of the evaluation. An evaluation is “empirically adequate” to the extent that it avoids these threats.⁷ In this section, we outline three conditions for empirical adequacy. To make our discussion more concrete, we use the example of measuring a system’s reliability. Since it is likely that the selected conception of epistemic performance includes a preoccupation for reliability – i.e., some weighing of the objectives of minimizing false claims (often called ‘precision’) and maximizing true claims (often called ‘sensitivity’ or ‘recall’) – our chosen example has the additional advantage of pointing to common difficulties with epistemic evaluations.

The first and most obvious condition for the empirical adequacy of an evaluation is:

Measurement Accuracy. The targeted properties must be accurately measured.

All other things being equal, we should favor an evaluation protocol for which we are confident that this condition holds.

The current degree of reliability of a system is often extremely hard to measure accurately in a *direct* manner. Indeed, measuring directly reliability implies that the epistemologist can discriminate what is true from what is false in the output of the system. In other words, the evaluator needs to be in some respects epistemically superior to the system to measure directly in an accurate way its current degree of reliability.

When reliability cannot be measured accurately in a direct manner, the epistemologist would be wiser to opt for an *indirect* strategy.⁸ This strategy is to measure factors that are thought to be positively correlated with what one seeks to determine – i.e., reliability in the present example. If the system is rich in these factors, the epistemologist can be more confident in its reliability. For instance, the internal social diversity of a system is often highlighted as contributing positively to the system’s

7 Our use of “empirical adequacy” is distinct from Van Fraassen’s use of the expression (Van Fraassen 1980) since we do not use the distinction between the observable and the unobservable. For us, empirical adequacy comes in degree and is achieved to the extent that the empirical part of the evaluation is successful, i.e., to the extent that it meets the three conditions specified in this section. We thank an anonymous referee for inciting us to make this distinction explicit.

8 See Dietsch, Claveau and Fontan (2018, 77–78) for the analogous choice, inspired by Longino’s (1990) procedural objectivity, of evaluating epistemic procedures instead of current output in the context of an appraisal of central bank expertise. Work on determining the conditions for justified consensus is another example of this indirect strategy (e.g., Miller 2013).

epistemic performance, and to its reliability in particular. Teams with diverse sociocultural and economic backgrounds and with wide expertise would tend to outperform more homogeneous teams (Page 2007; Intemann 2009). Note that diversity is thought to be a cause of reliability, but the indirect strategy can use factors that are correlated for other reasons with reliability.⁹

The condition of Measurement Accuracy is not sufficient for the empirical adequacy of any indirect measurement of performance – be it measuring diversity as a proxy for reliability or measuring another factor meant to be linked to epistemic performance. A further condition must be met:

Applicability of the Generalization. The generalization connecting the measured factor with epistemic performance is true of the system under study.

Indeed, if it is false that ‘the measured factor positively correlates with epistemic performance for the studied system’, the indirect route is broken. Obviously, the evaluator never knows for sure the real scope of a generalization, but we should favor, all other things being equal, an evaluation protocol relying on generalizations in which we are confident.

If the first two conditions are met, an indirect measurement of performance is empirically adequate in a *minimal* sense: after the measurement, the epistemologist can be more confident about the epistemic performance of the system, but the warranted increase in confidence might be mild. In particular, if the factor(s) measured account for only a small fraction of the variability in epistemic performance, the warranted conclusion will be weak. For example, measuring low diversity for a system can warrant a negative conclusion of the sort ‘this system fails to have one property contributing to epistemic performance.’ Yet, it will be hasty to conclude that this system underperforms epistemically since it is plausible that other (unmeasured) properties counterbalance the low diversity.

These considerations can be captured by our third condition, which is necessary for indirect measurements to be empirically adequate in a *maximal* sense:

Exhaustiveness of the measured factors. The factors measured account together for all the possible variation in epistemic performance.

Again, the epistemologist can never be certain that this condition is met. It serves as a guiding ideal: all other things being equal, the more the evaluation protocol measures factors that are thought to account for a large part of the variability in epistemic performance, the better it is for the empirical adequacy of the exercise.

To sum up, our goal in this subsection was to delineate three conditions contributing to the empirical adequacy of a measurement. Depending on the empirical evidence used in the evaluation – i.e., whether it comes from direct measurement of performance or not – the last two conditions may not always be

9 For instance, under the assumption that the past is a good guide to the present, the past reliability of a system – i.e., its track record of epistemic successes and failures – can be used as a proxy for its current degree of reliability. Strictly speaking, past reliability is not a cause of current reliability. If they are correlated, it is because they are effects of common causes (e.g., virtuous stable dynamics of the system).

relevant, but they must be kept in mind because the condition of Measurement Accuracy is typically not sufficient for empirical adequacy.

3.3 Practical relevance of the evaluation

An epistemic evaluation is typically motivated by the goal of improving practices. Indeed, epistemic evaluations are rarely done out of pure intellectual curiosity. Borrowing an analogy from Bishop and Trout (2017, 103), epistemologists typically think of themselves as akin to coaches who are tasked with counseling agents in order to ameliorate their epistemic performance. In consequence, whether an epistemic evaluation is practically relevant does much to justify the resources invested in producing it.

From an ameliorative perspective, an epistemic evaluation of any system can be useful in two ways:

1. It can influence the evaluated system to conform to the chosen conception of epistemic performance.
2. It can allow the other systems relying upon the evaluated system to make better informed choices.

The first type of desired change is probably the most obvious: the epistemologist acts as a coach for the evaluated system (or for components of the system), nudging the system toward a better performance. The second type of change stems from the fact that systems exist in networks of epistemic dependence. This is a truism for individual agents: we each take other individuals as sources for our beliefs (Hardwig 1985). This dependence is not limited to networks of individuals. For instance, organizations are epistemic sources for individuals and for other organizations. If a particular system is in a relation of epistemic dependence with another system, it can use the results of an epistemic evaluation to calibrate the level of trust it is willing to grant to this source.

These two uses of epistemic evaluation correspond to two conditions. At least, one condition must be met in order for the evaluation to be practically relevant.

Responsiveness of the Evaluated System. The evaluated system is likely to change or consolidate its practices in response to the results of the epistemic evaluation.

Responsiveness of the Dependent Systems. The systems which depend on the evaluated system as an epistemic source are likely to change or consolidate their practices in response to the results of the epistemic evaluation.

The actualization of these two conditions is not necessarily explained by the system's intrinsic motivation to be a better epistemic agent. First, the motivation can be extrinsic: the incentive structure might nudge the system toward epistemic performance even though it is not a goal of a system. Second, systems (e.g., networks or ecosystems) need not have motivation at all. Their responsiveness might come from changes in the incentive structure faced by agents that are part of the system. The source of the responsiveness is unimportant. What matters is that the evaluated system as well as the dependent

systems are responsive to evaluation and will modify their practices in predictable ways following a positive or negative epistemic evaluation.

4 Epistemic evaluations of think tanks

4.1 Our sample of evaluations

There is a large number of evaluations of think tanks, each of them focusing on different criteria. These evaluations are rarely explicitly epistemic. However, when considered from an epistemologist's point of view, underlying epistemic considerations can be attributed to most evaluations. That being said, the only common factor across evaluations is the explicit objective to rank think tanks from best to worst or to nominate some think tanks as 'best.' In so doing, all explicit evaluations to date place themselves at the organizational level. Beyond this common objective, there is considerable variety in the criteria and the methods used.

Our sampling strategy of existing evaluations has been to intentionally select instances that do not share criteria instead of embarking on the elusive quest of having an exhaustive list of instances. We thus focus on four evaluations that are, as far as we know, representative of the existing diversity of think tank evaluations: Transparify's ranking,¹⁰ Clark and Roodman's research,¹¹ the Atlas Network's prize,¹² and James McGann's ranking. The remainder of this section introduces each instance while Table 1 synthesizes some important differences.

Table 1: Sample of Think Tank Evaluations at the Organizational Level

Evaluator	Criterion	Method of evaluation
Transparify	Financial transparency	Qualitative
Clark and Roodman	Public attention	Quantitative
Atlas Network	Contribution to the promotion of free market	Qualitative
McGann	Multifaceted	Expert based ranking

First, the UK based Transparify uses transparency about funding as its only evaluation criterion. As long as funding is declared by a think tank, Transparify does not judge the source of the funding itself. The method used to rank think tanks is simple. Two independent raters evaluate the think tank's transparency before an adjudicator reviews the two ratings. In order to rate a think tank's financial transparency from "deceptive" to "five stars", it uses only the information that is readily available on

10 For another example of this type of evaluation, see the UK initiative "Who Funds You?", <http://whofundyou.org/> (last accessed: 2019-06-18).

11 For other examples of this type of evaluation, see (Ruble 2000; Posen 2002; Trimboth 2005)

12 For another example of this type of evaluation, see the (discontinued) ICCG Climate Think Tank Ranking, https://web.archive.org/web/20180703202342/http://www.thinktankmap.org/Page.aspx?Name=About_the_Ranking (archive last accessed: 2019-06-19; the website was not in operation anymore in June 2019).

the think tank's website. This type of evaluation is situated at the organizational level: it judges a think tank's financial transparency, staying focused on properties of the organization itself.¹³

Second, Clark and Roodman's evaluation focuses on the public attention received by a think tank – what they call its “public profile,” which should not be confused with influence (Clark and Roodman 2013, 3). To measure public attention, they use multiple factors related to various types of citation counts. One public being academia, they gather academic citation counts using Google Scholar in combination with the Publish or Perish software (Clark and Roodman 2013, 8). They combine these academic citations with the broader public attention of a think tank, which is measured by engagement with its platform on social media (Clark and Roodman 2013, 5) as well as its references in news media (Clark and Roodman 2013, 7) Evaluations based on public attention such as Clark and Roodman's focus on properties of particular think tanks and, as such, are situated at the organizational level.

Third, the Atlas Network's evaluation is based on a specific ideological criterion. The Atlas Network connects more than 450 think tanks all over the world and aims to strengthen the worldwide freedom (*read : free market*) movement.¹⁴ The Templeton Freedom Award is given out yearly to the think tank within the network that has made the most impactful and innovative contribution to free enterprise and free competition research and public policy. This type of evaluation is situated at the organizational level. While it samples from a network of think tanks, it focuses on properties which are tied to particular think tanks in order to rank them and honor the best among them with the Templeton Award.¹⁵

Fourth, James McGann's evaluation is the best known think tank ranking: the Go to Global Think Tank Index. To produce his ranking, McGann does not rely on one specific criterion. While, McGann does suggest the use of twenty-eight different criteria and four impact indicators, their use remains optional (McGann 2017, 21). Instead, the Go to Global Think Tank Index relies heavily on expertise. This expert-based ranking system, uses the various experts' criteria of choice. This type of evaluation is situated at the organizational level because it rates each think tank based on its properties.

4.2 Epistemic performance and think tanks

An epistemic evaluation always assumes a conception of epistemic performance. Since existing evaluations of think tanks are *implicitly* epistemic, the associated conception of epistemic performance is also implicit. In the following section, we would like to suggest guidelines for an explicit conception of epistemic performance which would be appropriate for an evaluation of think tanks. We contend that a purist conception of epistemic performance is not appropriate to evaluate think tanks. Veritism as proposed by Goldman misses important elements of epistemic performance when dealing with think tanks. An extended conception of epistemic performance would be a much better choice. This is based on the idea that both significance and what we will call ‘reach’ must be integrated into the conception of epistemic performance for think tanks. This is not to say that the resulting evaluation must give a

13 Website: <https://www.transparify.org/> (last accessed: 2019-06-18)

14 Website: <https://www.atlasnetwork.org/about/our-story> (last accessed: 2019-07-15)

15 Website: <https://www.atlasnetwork.org/grants-awards/awards> (last accessed: 2019-06-18)

unique global epistemic score to each think tank. There is little value in summarizing epistemic performance in a single number or qualifier since it does not tell us the relative strengths and weaknesses of each system. Notably, our general framework distances itself from existing think tank evaluations by not requiring a unitary evaluation.¹⁶ Now back to the explanation of the two components which make up our extended conception of epistemic performance: significance and reach.

First, significance must be taken into account. As we have seen above, purist conceptions of epistemic performance tend to sideline questions of interest. This is the case in Goldman's framework where questions of interest are assigned a minimal role: the questions answered correctly must only be of *some* interest to be fully counted in the estimation of epistemic performance (see section 3.1). A think tank working on a minor subject will thus be evaluated more positively than a think tank working on major subjects if the former manages to be less frequently wrong and more frequently right than the latter. This outcome is in fact likely given that more pressing questions – e.g., questions relevant to the survival of humanity – tend to involve more complexity and uncertainty. This result clashes with what we should expect of think tanks as contributors to collective knowledge seeking. Think tanks position themselves as actors who focus most of their research efforts on providing solutions to the most pressing problems faced by our societies. For instance, the C.D. Howe Institute states that its “research aims at understanding and providing options to address four key challenges central to Canadians’ prosperity”¹⁷. To incorporate the significance of topics in an evaluation of think tanks, we need an extended conception of epistemic performance.

Second, another factor lacking in a purist conception of epistemic performance is reach – i.e., the extent to which an organization's output is heard and taken into account by other agents. In the case of think tanks, output refers to everything from a think tank's official tweets to its scholarly publications. We contend that reach is particularly important in the case of the epistemic evaluation of think tanks because a think tank's epistemic states have no value in themselves. A think tank's epistemic state can only be of instrumental value through its impact on the epistemic states of individual agents. Consequentially, a think tank which would not have any reach would not modify agents' epistemic states and would therefore not be an epistemically relevant object of study.

The reach of a think tank's message varies on two dimensions. There is the extensive margin, which is simply the number of agents reached by the think tank's output. These agents can be journalists, policy-makers, academics or simple citizens. Then there is the intensive margin, which is the degree of engagement that the reached agents have with the think tank's output. Here, a case of high intensity engagement would be a causal chain between a think tank's output, a change in belief of policy makers and a policy change. A low intensity engagement would be a retweet of a think tank's message (which does not even imply a change of belief).

Variation on the intensity margin illustrates that reach is importantly different from influence. First, real influence at least implies changes in belief, and is often meant as changes in actions such as enacting a

16 We thank an anonymous reviewer for bringing this much needed precision to our attention.

17 Website: <https://www.cdhowe.org/objectives> (last accessed: 2019-06-18)

new policy. Reach does not require anything as stringent. There is a noteworthy similarity here with Clark and Roodman's focus on public attention: they too note that attention is not impact, although "ideas need to be noticed to be adopted" (Clark and Roodman 2013, 3).¹⁸

Second, existing evaluations of think tanks, including Clark and Roodman's, assume influence to be always a good thing: the more a think tank has influence, the better is its performance. Reach does not have this unambiguous relationship with *epistemic* performance. For instance, if reach is coupled with low reliability, high reach – especially on the intensive margin – makes for epistemically undesirable results. On the contrary, if reach is coupled with high reliability, high reach is epistemically advantageous. Therefore, reach must be part of an acceptable conception of epistemic performance for think tanks, which interacts with other considerations such as reliability and significance.

That being said, to our knowledge, this conception of epistemic performance is absent from think tank evaluations. Moreover, not only is our suggested conception of epistemic performance absent from the literature on think tank evaluation, no *explicit* characterization of epistemic performance is present. The conception of epistemic performance underlying the evaluation is always implicit. McGann and the *Go to Global Think Tank Index*, since it is the best known think tank ranking (Clark and Roodman 2013, 2), can serve as an emblematic case.

How does the ranking system of the Go to Global Think Tank Index work? The first step is extensive research to update the think tank database. This step is followed by the nomination of a panel of experts who then issue a call for nominations to think tanks. In 2016, the call for nomination was sent out to approximately 6800 think tanks and 4700 journalists, public and private donors and policy-makers. Think tanks that have ten nominations (or more) as well as the top think tanks from the previous year's rankings are allowed to be added to the ballot (McGann 2017, 5). Once this is done, a first round of expert ranking is carried out. For the last round of ranking, information packages are sent to the experts to help them make their final decision. These packages contain twenty-eight criteria and four indicators of impact, which experts are advised to use when making their decisions (McGann 2017, 21).

The Go to Global Think Tank Index's explicit goal is to "increase the profile, capacity and performance of think tanks at the national, regional and global levels so they can better serve policy makers and the public" (McGann 2017, 5). Note that "increasing performance" is explicitly listed as an objective. However, the conception of performance which underlies this goal is never made explicit. While a list of twenty-eight criteria is given, it is difficult to infer the underlying conception of performance (epistemic or not) from the list of essentially very different criteria. They include the "ability to recruit and retain elite scholars and analysts," (McGann 2017, 21) the "ability to use electronic, print and the new media to communicate research and reach key audiences" (McGann 2017, 22) and the "ability to bridge the gap between policymakers and the public" (McGann 2017, 23). Because the criteria cover a wide range of factors, it is difficult to piece together a coherent conception of epistemic performance.

18 A dissimilarity is that Clark and Roodman (2013, 3) explicitly leave aside reach "behind the scenes" – i.e., attention think tanks get from policy-makers that is not publicly visible.

That being said, there seems to be a pronounced emphasis on impact as evidenced by the provision of “four indicators of impact” which are given to the experts in addition to the twenty-eight criteria. Impact, according to McGann, is positive if it “changes the behaviour, relationships, activities, or actions of the people, groups, and organizations with whom a program works directly” (McGann 2017, 24). This is problematic from an epistemic standpoint. For instance, if a think tank was successful in convincing a large portion of the population that vaccination is dangerous, it would modify behavior (people would stop getting vaccinated). While, according to McGann, this should be registered as a positive instance of impact, it seems obvious that such an impact would be considered to be an epistemically worrisome outcome. Impact then is not always epistemically positive and, because the conception of performance is left implicit, little is done to justify the seemingly central role of impact.

Furthermore, the experts who receive the non compulsory list of twenty-eight criteria and four indicators are not told how to operationalize or weigh them. The weight given to each criterion thus depends on the individual expert’s conception of (epistemic) performance, making for an uneven evaluation. By having *numerous* experts rank think tanks based on a list of criteria and indicators that is long, ambiguous and non compulsory, it is probable that the final rankings are based on *incompatible* conceptions of performance. The end result is that agents consulting the rankings do not know why a particular think tank is ranked above another.

In summary, Patrick Koellner does a nice job succinctly expressing the issue with using such implicit conceptions of epistemic performance to evaluate think tanks by stating that “while such ranking indexes help to draw attention to the growing think tank scenes across the globe and are thus to be welcomed, the existing rankings are fraught with problems; conceptual and methodological difficulties in particular [...] abound” (Koellner 2013, 1).

4.3 Empirical adequacy of existing organizational evaluations

The existing literature on the evaluations of think tanks focuses on the organizational level. The concentration of evaluations at this level might indicate that it is the best choice when dealing with think tanks. We will argue otherwise. In what follows, we assess existing evaluations based on the conditions for empirical adequacy. To do so, we will test the four organizational evaluations of think tanks we presented previously (see Table 1) against our three conditions: *measurement accuracy*, *applicability of the generalization*, and *exhaustiveness of the measured factors* (see section 3.2 for details).

The first condition is measurement accuracy. Two of the four evaluations are problematic from the perspective of this condition: the properties on which the Atlas Network and the Go to Global Think Tank Index focus are unclear. They both seem to be after an ‘impact’ of some sort. Yet, the sort of impact and the factors used to measure this property are opaque to outside observers. It is thus difficult to assess whether the properties are accurately measured. The two remaining evaluations, which are Transparify’s and Clark and Roodman’s, fair better. They have clear protocols to measure their property of choice. Transparify measures the accessibility of the funding information on the think tank’s website,

and its protocol with two raters and an adjudicator is designed for accuracy. Clark and Roodman measure citations in academic journals and in mass media, and describe quite precisely their protocol such that anyone could reproduce their results.

The second empirical adequacy condition is the applicability of the generalization. We need to supply some interpretation here because, as we have noted, no evaluation incorporates an explicit conception of epistemic performance, meaning that no evaluation connects explicitly through a generalization what it measures with better or worse epistemic performance. We change the order of presentation of the evaluations here to start with cases for which a plausible generalization comes more readily to mind.

In the case of Transparify, focusing on financial transparency can be justified based on the generalization that ‘A more financially transparent think tank will be more reliable.’ Transparency about conflicts of interest is a well-established practice in other epistemic systems. The identification of a conflict of interest is sometimes judged to be sufficient ground to exclude an agent from the epistemic process – e.g., in the jury system. In other cases, disclosure of the conflict of interest is taken to be sufficient – e.g., in the academic publication system. In the latter cases, it is expected that agents disclosing the conflict of interest will adopt more reliable epistemic practices because a seemingly erroneous method, reasoning or result will be readily attributable by other agents to the presence of this conflict. Is this expectation warranted for think tanks? If it is, the generalization would be applicable to the system under study (as our second condition requires). Without fully answering the question, we can say, at least, that this generalization seems to us more secure than the ones that could justify the other evaluations.

In the case of Clark and Roodman’s evaluation, measuring public attention can be interpreted as a *direct* strategy to determine one aspect of an extended conception of epistemic performance: reach.¹⁹ Generalizations are not needed for direct strategies. Yet, there is a more ambitious interpretation of Clark and Roodman’s evaluation: public attention could be taken as indicative of other aspects of epistemic performance such as reliability and significance. The underlying generalization would be: ‘Think tanks garner more public attention because they are reliable and produce information on significant topics.’ This generalization is not without grounds outside the field of think tanks. In academia for instance, the high citation count of a scholarly article is an indication that many researchers have noticed it, but also that it is on a significant topic for many researchers and that it is generally taken to be reliable. However, the generalization does not travel well to the field of think tanks, especially when public attention is taken to be indicative of reliability: agents engaging with the contents of think tanks often do so for entirely other reasons. Clark and Roodman (2013, 20) admit this limitation. After highlighting that the Heritage Foundation and the Cato Institute lead their rankings, they state: “One possible explanation for these extreme outliers could be that many people who follow these and other more ‘ideologically driven’ tanks on social networks do so in part as a values statement.” As long as this explanation is plausible, measuring public attention can only be indicative of reach, not reliability. Since reach, by itself, does not say much about epistemic performance –

19 Under this interpretation, one worry related to our first condition for empirical adequacy is that it misses the reach ‘behind the scenes’, see footnote .

remember that high reach for an unreliable source is an epistemic liability (see section 4.2) – measuring public attention does not carry us far in our quest for an epistemic evaluation.

Since the last two evaluations in our sample are unclear about what factors they intend to measure, we cannot even begin to interpret which generalizations would establish that these factors are indicative of epistemic performance. However, they seem to be each working with a generalization that is highly problematic from the point of view of epistemology. The Atlas Network seems to assume that the results of research are predetermined: good research is research that highlights the benefits of “free competition” and convinces countries to improve their “scores in ranking of economic freedom”.²⁰ The possibility that a piece of research doing exactly the opposite could be epistemically better is not entertained. The Go to Global Think Tank Index seems to assume that its experts know what to evaluate and how to evaluate it. But it is again likely that it just aggregates different views of what is a ‘good think tank’, turning the whole enterprise into a popularity contest.

The third empirical adequacy condition is the exhaustiveness of the measured factors. All four evaluations struggle with this final condition because they all take place at the organizational level. They thus miss factors that are epistemically salient, but situated at the level of the network or the ecosystem.

To illustrate this point, we can use the example of the level of “public attention” (Clark and Roodman 2013, 3). If think tanks were academic research teams publishing scientific articles, we could justifiably use the level of academic attention of their research as an indicator of epistemic performance. This empirical protocol would be justifiable because of a property of the ecosystem in which academic research teams operate: the vigilance of other members, or what Robert Merton (1942, 126) called the “organized skepticism” of science. Although the norm is not always followed, “the detached scrutiny of beliefs in terms of empirical and logical criteria” (Merton 1942, 126) is highly valued in the academic ecosystem. In contrast, the level of vigilance in the think tank ecosystem can vary substantially. For instance, professional journalists might serve as gatekeepers for the general public by filtering the transmission of a think tank’s messages based on an appraisal of its reliability. If this property of the ecosystem changes – either by a relaxation of journalistic standards or by the creation of social media that bypass journalists – the epistemic import of high public attention is transformed.

The same point could be made with other factors at the organizational level. For instance, funding transparency is likely to significantly affect reliability only if think tanks are worried that vigilant agents will not accept shaky research designs because they now know who funds the research. In short, an evaluation focusing on organizational factors at the exclusion of ecosystemic factors is unlikely to account for most of the variation in epistemic performance. In other words, organizational factors are clearly far from exhausting the factors relevant to this type of variability.

20 Website: <https://www.atlasnetwork.org/grants-awards/awards> (last accessed: 2019-06-18)

Table 2: Summary of results about the empirical adequacy of the evaluations

	Transparify	Clark and Roodman	Atlas Network	Go to Global Think Tank Index
Criteria	Financial transparency	Public attention	Contribution to the promotion of free market	Multifaceted
Measured factor	Accessibility of the funding information on the think tank's website	Citation in academic journals and mass media	Unknown	Unknown
Measurement issues	Few	Few	Worrisome	Worrisome
Direct/Indirect	Indirect	Direct/Indirect	Indirect	Indirect
Generalization	A more transparent think tank is more reliable	Think tanks garner more public attention because they are reliable and produce information on significant topics	Good research is research that highlights the benefits of "free competition"	Unknown
Applicability of generalization	Barely acceptable	Somewhat	Worrisome	Inapplicable
Exhaustiveness	Worrisome	Worrisome	Worrisome	Worrisome

Table 2 sums up the results of this section on the empirical adequacy of our sample of evaluations. We have seen that whether an evaluation meets the first condition of measurement accuracy is contingent in large part upon the evaluators' choice of measured factors. The issues raised regarding the two other conditions – the applicability of the generalization and the exhaustiveness of the measured factors – are more general problems stemming from the decision to remain at the organizational level.

4.4 Relevance of existing evaluations

To be relevant, an organizational evaluation of think tanks should be able to modify the practices of the organization or should be able to modify the practices of other systems which rely on think tanks.

First, how might an epistemic organizational evaluation prompt the evaluated think tank to improve its epistemic practices? Organizations can modify their practices in the same way that individuals can modify their knowledge-seeking practices to conform to certain standards. If an organization is intrinsically motivated to excel epistemically, a negative evaluation can push it to modify its practices while a positive evaluation can comfort it in its habits. The evaluation gives such organizations the

necessary information to decide if adjustments should be made. Based on concerns for its reputation, an organization can also be extrinsically motivated to conform to the conception of epistemic performance put forward by an evaluation. In the case of think tanks, positive evaluations are often proudly displayed on the front page of official websites. On the other hand, negative evaluations can damage reputations and hurt credibility. Even if a think tank does not intrinsically care about being an excellent epistemic system, it might be in its advantage to take such bad evaluations to heart. By way of illustration, Transparify reports having witnessed a significant trend in think tanks leaning toward financial transparency after it started evaluating them on this ground (Gutbrod 2018, 3).

Second, how might an epistemic evaluation inform the decisions of agents who rely on the organization in question when enacting their own knowledge-seeking practices? In this case, even though the evaluation concerns the organization, its usefulness is derived from the way in which individuals will interpret it. In the case of think tanks, an external agent (e.g., a journalist, a bureaucrat or an ordinary citizen) might become more skeptical of a think tank's claims upon learning that this think tank was negatively evaluated. Of course, the opposite experience is also possible. Upon learning that a think tank has been positively evaluated, an external agent might consider the think tank's claims with less suspicion. For instance, the Montreal Economic Institute was rated as highly opaque by Transparify (Gutbrod 2017, 6). This might lead agents to modify their degree of trust in the think tank's publications.

That being said, there are reasons to doubt that the two conditions associated with these ameliorative functions are frequently fulfilled by the existing evaluations of think tanks. We pinpoint weaknesses of epistemic organizational evaluation that suggest that another level of epistemic evaluation might be a better choice to study these particular objects if one wishes to fulfill the relevance conditions.

The satisfaction of the first condition, which consists in the responsiveness of the evaluated system, is impeded by the fact that a think tank's practices are mainly determined by higher level forces (Medvetz 2012). A think tank has very little room to change without risking an unraveling of its specific ties with actors from neighboring fields. Without further changes in the ecosystem, the pressure against reform emanating from the other forces at work will be high. An epistemic organizational evaluation of think tanks does not take this into account and, when an evaluation ignores the balancing act a think tank must perform between different fields in order to thrive, its potential for reform is reduced significantly. Furthermore, because think tanks react to demands that stem from complex interactions, if a think tank simply changes its identity to comply with certain epistemic standards, it is highly likely that another think tank will rise up and fill the newly vacated niche (Landry 2018, 126). Knowing this, compliance becomes an unappealing option which in turn reduces the evaluation's potential for reform.

If this is true and little change can be expected from organizational evaluations, why has Transparify reported an increase in transparency? First and foremost, the attribution of a causal chain between Transparify's evaluations and increased overall transparency in think tanks is not something which has been solidly established. The increase in transparency might be caused by other factors. Furthermore, it is possible that most think tanks will see in transparency a net gain of symbolic capital (or, in reverse, a

risk of losing symbolic capital if they do not comply) while still being able to cater to the interests of actors in other fields (e.g., funders, political parties).

The satisfaction of the second condition, which consists in the responsiveness of the dependent systems, is impeded because an organizational evaluation shifts the bulk of the epistemic labor onto individual agents. To serve as guides, evaluations need to be actively sought out. As such, only highly motivated agents will do the work that this system of evaluation requires of them when they are in search of information. This seems like an excessive burden to place on an agent who must already fight against motivated reasoning in her search for knowledge. Moreover, because of the diversity of organizational evaluations that exist, it is easy for an agent to find an evaluation that comforts her initial decision to trust one think tank over others and avoid evaluations which challenge her initial impression. For instance, an organizational evaluation such as Transparify's forces individual agents to look up the transparency score of each specific think tank. Even more labor intensive, it forces agents to look up different evaluations and understand the specificities of each in order to adjust their level of trust accordingly.

5 Conclusion

The primary function of think tanks should be to produce and disseminate knowledge relevant to public policy. This is how they can serve society. An epistemic evaluation of think tanks aims to determine whether think tanks serve this function well.

This article's aim is twofold. It is a first step in building a solid framework for epistemic evaluation of social systems. The literature in social epistemology is lacking so far in this regard, especially when it comes to discussing the empirical dimension of epistemic evaluations. The article also paves the way for further work on the evaluation of think tanks by assessing existing evaluations – i.e., a meta-evaluation. As a necessary step in a rigorous meta-evaluation, we have applied our proposed conceptual framework to four representative evaluations of think tanks and have, as a result identified serious limitations with the existing work. In this conclusion, we want to highlight two general issues with existing evaluations.

First, many evaluations blur the line between the primary societal function of think tanks – i.e., producing and disseminating knowledge on public policy – and the functions attributed to think tanks by their funders and other interested parties. There is no doubt that some agents have non-epistemic interests that think tanks can serve in a better or worse way: think tanks can be powerful tools in power struggles. When an evaluation focuses on how far a think tank's message reaches or how influential its research is, it does not properly distinguish between the societal function and the political functions it can serve. An explicitly epistemic evaluation should do a better job distinguishing between the two vastly different functions.

Second, all existing evaluations of think tanks take place at the organizational level: their aim is to rate each think tank and thus highlight the 'best' in the lot. If our goal is to improve the global epistemic

performance of think tanks, this choice of level has serious drawbacks. Most importantly, organizational evaluations miss factors that are situated at the network and at the ecosystemic levels and that significantly determine how well think tanks serve their epistemic function. The literature on think tanks in sociology and political science has highlighted how dependent think tanks are on other fields (Medvetz 2012; Abelson 2016). The ecosystem of think tanks includes other think tanks, but also organizations from the academic, the media, the financial, the political and the bureaucratic fields. How these fields relate to think tanks – for instance, how vigilant they are about the reliability of their research – is crucial to the latter’s epistemic performance. Since existing evaluations do not take this fact into account, there is a need for developing an ecosystemic evaluation of think tanks.

References

- Abelson, Donald E. 2000. “Do Think Tanks Matter? Opportunities, Constraints and Incentives for Think Tanks in Canada and the United States.” *Global Society* 14 (2): 213–36. <https://doi.org/10.1080/13600820050008458>.
- . 2016. *Northern Lights: Exploring Canada’s Think Tank Landscape*. Montréal: McGill-Queen’s University Press.
- Bishop, Michael A., and J.D. Trout. 2005. *Epistemology and the Psychology of Human Judgment*. Oxford: Oxford University Press.
- . 2008. “Strategic Reliabilism: A Naturalistic Approach to Epistemology.” *Philosophy Compass* 3/5: 1049–65.
- Bishop, Michael, and J.D. Trout. 2017. “Epistemology for (Real) People.” In *A Companion to Applied Philosophy*, edited by Kasper Lippert-Rasmussen, Kimberley Brownlee, and David Coady, 103–19. Chichester, UK: John Wiley & Sons.
- Bluhm, Robyn, and Kirstin Borgerson. 2019. “Evidence-Based Medicine.” In *The Routledge Handbook of Applied Epistemology*, edited by David Coady and James Chase. Routledge Handbooks in Philosophy. Abingdon, Oxon: Routledge.
- Clark, Julia, and David Roodman. 2013. “Measuring Think Tank Performance: An Index of Public Profile.” Center For Global Development. June 2013.
- Dietsch, Peter, François Claveau, and Clément Fontan. 2018. *Do Central Banks Serve the People?* Cambridge, UK: Polity Press.
- Drezner, Daniel. 2017. *The Ideas Industry: How Pessimists, Partisans, and Plutocrats Are Transforming the Marketplace of Ideas*. New York, NY: Oxford University Press.
- Fallis, Don. 2008. “Toward an Epistemology of Wikipedia.” *Journal of the American Society for Information Science and Technology* 59 (10): 1662–74. <https://doi.org/10.1002/asi.20870>.
- Fraassen, Bas C. van. 1980. *The Scientific Image*. Oxford: Clarendon Press.
- Gilbert, Margaret. 2014. *Joint Commitment: How We Make the Social World*. Oxford: Oxford University Press.
- Goldman, Alvin I. 1999. *Knowledge in a Social World*. Oxford: Oxford University Press.
- . 2000. “Replies to Reviews of *Knowledge in a Social World*.” *Social Epistemology* 14 (4): 317–33. <https://doi.org/10.1080/02691720010008662>.

- . 2002. “Reply to Commentators.” *Philosophy and Phenomenological Research* 64 (1): 215–27. <https://doi.org/10.1111/j.1933-1592.2002.tb00155.x>.
- . 2010. “Systems-Oriented Social Epistemology.” In *Oxford Studies in Epistemology*, edited by John Hawthorne and Tamar Szabó Gendler, 3:288.
- Gutbrod, Hans. 2017. “Think Tank Transparency in Canada : Lagging behind the US and UK.”
- . 2018. “Transparify 2018 Report.” *Transparify*. July 17, 2018.
- Hardwig, John. 1985. “Epistemic Dependence.” *The Journal of Philosophy* 82 (7): 335–49. <https://doi.org/10.2307/2026523>.
- Hodgson, Geoffrey M. 2006. “What Are Institutions?” *Journal of Economic Issues* 40 (1): 1–25.
- Intemann, Kristen. 2009. “Why Diversity Matters: Understanding and Applying the Diversity Component of the National Science Foundation’s Broader Impacts Criterion.” *Social Epistemology* 23 (3–4): 249–66. <https://doi.org/10.1080/02691720903364134>.
- Kinzel, Katherina. 2015. “Narrative and Evidence. How Can Case Studies from the History of Science Support Claims in the Philosophy of Science?” *Studies in History and Philosophy of Science Part A* 49 (February): 48–57. <https://doi.org/10.1016/j.shpsa.2014.12.001>.
- Koellner, Patrick. 2013. “Think Tanks: The Quest to Define and to Rank Them.” *GIGA Focus International Edition English*, no. 10.
- Kotzee, Ben. 2013. *Education and the Growth of Knowledge: Perspectives from Social and Virtue Epistemology*. John Wiley & Sons.
- Landry, Julien. 2018. “Les Think Tanks.” In *Experts, Sciences et Sociétés*, edited by François Claveau and Julien Prud’homme. Montréal: Presses de l’Université de Montréal.
- Laudan, Larry. 2006. *Truth, Error, and Criminal Law: An Essay in Legal Epistemology*. Cambridge: Cambridge University Press.
- Lindquist, Evert. 2003. “Think Tanks and the Ecology of Public Policy.” In *Banking on Knowledge: The Genesis of the Global Development Network*. Routledge.
- Longino, Helen E. 1990. *Science as Social Knowledge*. Princeton: Princeton University Press.
- McAllister, James W. 2018. “Using History as Evidence in Philosophy of Science: A Methodological Critique.” *Journal of the Philosophy of History* 12 (2): 239–58. <https://doi.org/10.1163/18722636-12341384>.
- McGann, James. 2007. *Think Tanks and Policy Advice in the US : Academics, Advisors and Advocates*. Routledge. <https://doi.org/10.4324/9780203963203>.
- . 2017. “2016 Global Go To Think Tank Index Report.” *TTCSP Global Go To Think Tank Index Reports*, January.
- Medvetz, Thomas. 2012. *Think Tanks in America*. Chicago: University of Chicago Press.
- Merton, Robert K. 1942. “A Note on Science and Democracy.” *Journal of Legal and Political Sociology* 1: 115–26.
- Miller, Boaz. 2013. “When Is Consensus Knowledge Based? Distinguishing Shared Knowledge from Mere Agreement.” *Synthese* 190 (7): 1293–1316.
- Page, Scott E. 2007. *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies*. Princeton, N.J.: Princeton University Press.
- Posen, Adam S. 2002. “Think Tanks: Who’s Hot and Who’s Not.” *The International Economy*, 2002.
- Ruble, Nicolas S. 2000. “Think Tanks: Who’s Hot and Who’s Not .” *The International Economy* 14 (5): 10–16.
- Schickore, Jutta. 2018. “Explication Work for Science and Philosophy.” *Journal of the Philosophy of History*, February. <https://doi.org/10.1163/18722636-12341387>.

- Scholl, Raphael, and Tim Rüz. 2016. "Towards a Methodology for Integrated History and Philosophy of Science." In *The Philosophy of Historical Case Studies*, edited by Tilman Sauer and Raphael Scholl, 319:69–91. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-30229-4_5.
- Stone, Diane. 2003. *Banking on Knowledge: The Genesis of the Global Development Network*. Routledge.
- . 2007. "Recycling Bins, Garbage Cans or Think Tanks? Three Myths Regarding Policy Analysis Institutes." *Public Administration* 85 (2): 259–78. <https://doi.org/10.1111/j.1467-9299.2007.00649.x>.
- Trimbath, Susanne. 2005. "Think Tanks: Who's Hot and Who's Not.," 2005.
- Weaver, R. Kent. 1989. "The Changing World of Think Tanks." *PS: Political Science & Politics* 22 (3): 563–78. <https://doi.org/10.2307/419623>.
- Wray, K. Brad. 2001. "Collective Belief And Acceptance." *Synthese* 129 (3): 319–33. <https://doi.org/10.1023/A:1013148515033>.