Epistemic Evaluations of Think Tanks: A Meta-Evaluation

by Andréanne Veillette and François Claveau
Department of Philosophy and Applied Ethics, Université de Sherbrooke

Abstract

In the following article we elaborate a conceptual framework with which one can systematize the evaluation of socioepistemic systems. The proposed framework contains four components which are equally important. Consequently, we have argue that an adequate epistemic evaluation must choose a level at which it will evaluate its object, it must explicitly define its chosen conception of epistemic performance, it must meet three conditions for empirical adequacy and it must meet both conditions for practical relevance. We then demonstrate how the framework can be applied to four representative evaluations of think tanks, our socioepistemic system of choice. The results indicate that the chosen level of evaluation for think tanks, which is the organizational level, impedes the evaluations' potential to properly assess a think tank's epistemic potential. The issues raised suggest another avenue of evaluation better fitted to the assessment of think tanks: the ecosystemic evaluation.

1 Introduction

The interest in socioepistemic systems as an object of study is a recent development in social epistemology. Being recent, much work remains to be done in ways of systematizing the field. Alvin Goldman, the instigator of this type of epistemology, even goes so far as to call it the "least familiar and most adventurous form of [social epistemology]" (Goldman and Whitcomb 2011, 11). That being said, system-oriented social epistemology represents a rich and important area of study because its objects are, most notably, public institutions which produce and disseminate knowledge. These institutions 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. 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 (Landry 2018). 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). 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.

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 evaluation in a systematic way in order to find the areas where further efforts are needed. In what follows, we assess the existing literature on think tank 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 with which to assess evaluations of think tanks. This framework is comprised of four necessary components: the level at which the evaluation takes place, the chosen conception of epistemic performance, the evaluation's degree of 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 evaluation of evaluations – our 'meta-evaluation' if you will – indicates that the existing evaluations share a blindspot, i.e., their choice to evaluate think tanks in isolation from properties of their network and ecosystem.

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,[1] 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 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 regulate a network – for instance, a league for a network of sport teams – but this organization is not identical to the network.

1 In social epistemology, much work is done on "groups" as doxastic agents – i.e., collective agents having belief-like attitudes. Groups can be understood as a special type of systems given our definition, although Alvin Goldman (Goldman 2011) separates group epistemology from systems-oriented social epistemology.

Finally, we draw a distinction between a network of similar organizations and an ecosystem. 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, the political institution and funders, interacting in a given environment.

3 Assessing epistemic performance

An epistemic evaluation is a specific type of system assessment: 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 in evaluating epistemic evaluations. The first subsection is about the issue of selecting a justifiable conception of epistemic performance. The second subsection focuses on issues with the empirical basis of the assessment. The last subsection turns to concerns about the relevance of the evaluation.

3.1 Conceptions of Epistemic Performance

An epistemic evaluation is a type of assessment 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.

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 that 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's 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 "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 2016, 111).[2] 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? Our proposed answer to this question uses two principles that can be called nonfoundationalism and contextualism.

Our approach is nonfoundationalist in rejecting the existence of any infallible source of guidance in deciding on a conception of epistemic performance. There is neither a fact of the matter nor apodictic intuitions to build on. We all have strong intuitive judgments about what an excellent knower is like, but these intuitive judgments can be revised if they give rise to unacceptable implications once formalized. We thus submit that a conception of epistemic performance must arise through a process of reflective equilibrium between our judgments and alternative conceptions (Goodman 1955, 674; Rawls 1999, 42–43).

Our approach is contextualist in doubting that a unique conception of epistemic performance is viable for all systems. In particular, it seems unlikely that the factors constitutive of epistemic excellence for human individuals are exactly those constitutive of epistemic excellence for higher-level systems such as organizations, networks and ecosystems. For instance, the positive epistemic states of an individual human – e.g., true beliefs or knowledge – have value in themselves because knowing is arguably part of any appealing conception of the good life. However, the positive epistemic states of organizations

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

have only derived (or instrumental) value: the fact that an organization is committed to a true (or false) claim gains normative traction only if it affects individual agents.

In sum, selecting a conception of epistemic performance through a contextualized process of reflexive equilibrium is a preliminary step necessary for any rigorous epistemic evaluation.

3.2 Empirical adequacy of the assessment

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 empirical adequacy of the exercise. In this section, we outline three conditions for empirical adequacy. To make our discussion more concrete, we use the example of measuring the reliability of the system. 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 and maximizing true claims – 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 assessment 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 assessor 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. 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 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.[3]

3 For instance, under the assumption that the past is a good guide to the present, the current degree of reliability of a system can be proxied by its past reliability – i.e., its track record at epistemic successes and failures. Measuring track record can be easier than measuring current reliability because the success of predictions can be checked or past claims can be compared to current understanding. 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).

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 assessor 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 fully 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 assessment – i.e., whether it comes from direct measurement of performance or not – the last two conditions may not always be 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 assessment is typically motivated by the goal of improving practices. Indeed, epistemic assessments are rarely done out of pure intellectual curiosity. Borrowing an analogy from Bishop and Trout (2016, 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 an other system, it can use the results of an epistemic evaluation to calibrate the level of trust it is willing to grant to this system.

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 of information 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,[4] Clark and Roodman's research,[5] the Atlas Network's prize,[6] 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 U.K. 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 assess 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 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.[7]

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.

4 For another example of this type of evaluation, see the website: http://whofundsyou.org/ (last accessed: 2019-06-18)
5 For other examples of this type of evaluation, see (Ruble 2000; Posen 2002; Trimbath 2005)
6 For another example of this type of evaluation, see the website: 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)
7 Website: https://www.transparify.org/ (last accessed: 2019-06-18)

Third, the Atlas Network's evaluation is based on a specific ideological criterion. The Atlas Network connects more than 450 think tanks in nearly a hundred countries and aims to strengthen the worldwide freedom *(read : free market)* movement. 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.[8]

Fourth, James McGann's evaluation is one of the best known think tank rankings: the Go to Global Think Thank Index. To produce his rankings, 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 in order to create rankings. This type of evaluation is situated at the organizational level because it focuses on the individual think tank's properties in order to rate them.

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. An extended conception of epistemic performance would be a much better choice. This is based on the idea that both significance and reach must be integrated into the conception of epistemic performance which will be applied in an evaluation of think tanks.

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 the providing solutions to the most pressing problems faced by of 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

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

Canadians' prosperity"[9]. 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. As explained above (see section 3.1), reach is 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 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).[10]

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 tanks 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* being the best known think tank ranking (Clark and Roodman 2013, 2) can serve as an emblematic case.

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

10 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.

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.

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 bad, 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 impact must play.

Furthermore, the experts which 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 think tank.

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 are 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 at this level might indicate that it is the best choice when dealing with think tanks. We will argue otherwise. In what follows, we will 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 that the Atlas Network and the Go to Global Think Tank Index focus on 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 who would wish to 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 of 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.[11] 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 – 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 convince countries to improve their "scores in ranking of economic freedom".[12] 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 assess and how to assess 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

11 Under this interpretation, one worry related to our first condition for empirical adequacy is that it misses the reach 'behind the scenes', see footnote 10.

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

"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 assessment 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 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.

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 tanks 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 evaluator's choice of measured factor. The issues which were raised regarding the two other conditions – generalization applicability and exhaustiveness of the measured factors – are deeper problems which, if not already doomed by the lack of clear measured factors, find their source in 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 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 itself, 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).Put simply, because of its low level of autonomy, a think tank has very little room to change.

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. 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. For instance, 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 constraint 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 the other and avoid evaluations which challenge her initial decision to trust a think tank. For instance, an organizational evaluation such as Transparify's forces individual agents to look up the transparency score a specific think tank received. 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 assess whether think tanks serve this function well.

This article is a first step in building a solid epistemic evaluation of think tanks. Its main aim is to evaluate existing evaluations – i.e., a meta-evaluation. As a necessary step in a rigorous meta-evaluation, we have elaborated a conceptual framework (sections 2 and 3) about inevitable choices made when evaluating socioepistemic systems and about the conditions an evaluation should meet. By applying our framework to four representative evaluations of think tanks, we have identified serious limitations within the existing work. In this conclusion, we want to highlight two limitations.

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 of 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 need for developing an ecosystemic evaluation of think tanks.

References

Abelson, Donald E. 2016. *Northern Lights: Exploring Canada's Think Tank Landscape*. Montréal: McGill-Queen's University Press.

Bishop, Michael A., and J. D. Trout. 2008. "Strategic Reliabilism: A Naturalistic Approach to Epistemology." *Philosophy Compass* 3/5: 1049–65.

Bishop, Michael A., and J.D. Trout. 2005. *Epistemology and the Psychology of Human Judgment*. Oxford: Oxford University Press.

———. 2016. "Epistemology for (Real) People." In *A Companion to Applied Epistemology*. Chichester, West Sussex ; Hoboken, NJ: John Wiley & Sons.

Clark, Julia, and David Roodman. 2013. "Measuring Think Tank Performance: An Index of Public Profile." Center For Global Development. June 2013.

Fallis, Don. 2006. "Epistemic Value Theory and Social Epistemology." *Episteme* 2 (3): 177–88. https://doi.org/10.3366/epi.2005.2.3.177.

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.

———. 2011. "A Guide to Social Epistemology." In *Social Epistemology: Essential Readings*, edited by Alvin I. Goldman and Dennis Whitcomb, 11–37. New York: Oxford University Press.

Goldman, Alvin I., and Dennis Whitcomb, eds. 2011. *Social Epistemology: Essential Readings*. New York: Oxford University Press.

Goodman, Nelson. 1955. *Fact, Fiction, and Forecast*. Cambridge, MA: Harvard University Press.
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.
Koellner, Patrick. 2013. “Think Tanks: The Quest to Define and to Rank Them.” *GIGA Focus International Edition English*, no. 10.
McGann, James. 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.
Page, Scott E. 2007. *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies*. Princeton, N.J.: Princeton University Press. http://press.princeton.edu/titles/8757.html.
Posen, Adam S. 2002. “Think Tanks: Who’s Hot and Who’s Not.” *The International Economy*, 2002.
Rawls, John. 1999. *A Theory of Justice, Revised Edition*. Cambridge, Mass: Belknap Press of Harvard University Press.
Ruble, Nicolas S. 2000. “Think Tanks: Who’s Hot and Who’s Not .” 2000. https://search.proquest.com/openview/163f40286d959ad2eed8776d85cd6980/1?pq-origsite=gscholar&cbl=30294.
Trimbath, Susanne. 2005. “Think Tanks: Who’s Hot and Who’s Not.,” 2005.