DOI: 10.1017/psa.2023.9
This is a manuscript accepted for publication in *Philosophy of Science*.
This version may be subject to change during the production process.

Information Deprivation and Democratic Engagement

Adrian K. Yee

(University of Toronto, IHPST)

Abstract

There remains no consensus among social scientists as to how to measure and understand forms of information deprivation such as misinformation. Machine learning and statistical analyses of information deprivation typically contain problematic operationalizations which are too often biased towards epistemic elites' conceptions that can undermine their empirical adequacy. A mature science of information deprivation should include considerable citizen involvement that is sensitive to the value-ladenness of information quality and that doing so may improve the predictive and explanatory power of extant models.

Introduction

Machine learning algorithms and statistical analysis are increasingly used to model information deprivation, such as misinformation, malinformation, and disinformation, with many models' accuracy and precision scores claimed to be higher than 95% (Alenezi & Alqenaei 2021, 13). However, models in these studies often make under analyzed and controversial methodological assumptions worthy of further philosophical analysis, raising questions regarding their success conditions. The first is that information deprivation is to be understood relative to a concept of *objective truth* that can allow for normative judgments about the quality of information in a manner that is sufficiently divorced from confounding value judgments. The second is that information deprivation is measured with respect to a set of *epistemic elites*, which are either academic researchers, journalists, or users of key websites that function as 'fact checkers' that adjudicate information quality. Thirdly, there is trust in the *mechanical objectivity*[1] of algorithmically induced statistical analysis and that the construct validation procedures, beginning with human conceptions of information deprivation, are preserved at the end state where algorithms must make future predictions on their own.

This paper seeks to improve upon this methodological orthodoxy by arguing that information scientists should reorient their methods of assessing the quality of information away from focusing on truth and unwarranted deference to epistemic elites, and that democratic elements should be further incorporated into adjudicating the quality

[1]See Daston & Galison (2007) for more on the history of this concept.

https://doi.org/10.1017/psa.2023.9 Published online by Cambridge University Press

of information that are participatory, transparent, and fully negotiable by average citizens, at least in principle. I conclude that increasing participation from citizens may even enhance the predictive and explanatory adequacy of our models of information deprivation.

1 Measuring Information Deprivation

There are two major schools of thought in the social science of information deprivation: the *capabilities* approach and the *incidence* approach. The capabilities approach argues that information deprivation occurs whenever individuals severely lack either the raw physical informational infrastructure, such as libraries or internet connection, or lack the relevant concepts and hermeneutical resources to understand or obtain relevant information that can allow individuals to achieve basic goals in their life (Britz 2004). Not knowing how or where to obtain clean drinking water or lacking access to telecommunications infrastructure required to bring goods and services to competitive markets are two common examples of severe deprivation of informational capabilities in developing nations. Deprivation of hermeneutic resources in a society can include how a person of LGBTQ identity fails to be identified adequately by their self-identified social category in their broader society, which can threaten physical safety in countries such as Iran and Saudi Arabia. Sincerely believing that the earth is flat is a consequence of an epistemically impoverished worldview that is increasingly common in developed nations such as the United States. The tragedy of Tanzanian albinos who are hunted by witchcraft practitioners for albinos' body parts is a salient under studied example of severe information deprivation, with respect to a lack of appropriate hermeneutical resources to make sense of albinism in several sub-regions of Tanzania (Bryceson et al. 2010).

My focus here will be on the increasingly common incidence approach in which units of information, rather than the set of social or material conditions impacting agents, are assessed for their informational quality. Photographs can be informationally deficient because they are doctored or the perspective taken biases the viewer in a certain fashion; social media posts can contain misleading or inaccurate statements; video testimonies frame events in a manner that obscures their occurrence. Measurement is typically quantitative only at the level of counting individual instances of information deprivation. In what follows, I critically analyze core components of the epistemic supply chain governing the construction of algorithmically induced machine learning and statistical models as a case study illustrating how fact-checking websites and academic scholars function as a problematic set of epistemic elites which typically unilaterally dictate informational quality in a manner that threatens the construct validity of their models.

As a first and representative example of these methods, Guess et al. (2019) collect 1191 Facebook users' posting histories during the Trump administration period and drew the following conclusions: 90% of respondents report that to the best of their knowledge they have not shared fake news; Republicans were more likely than Democrats to share fake news (38 vs 17 respective respondents); seniors over the age of 65 were most likely to share fake news to their Facebook friends, even holding constant education, ideology, and

https://doi.org/10.1017/psa.2023.9 Published online by Cambridge University Press

partisanship; there is no strong correlation between those who share lots of news items and the proportion of those items being fake news items; and those with conservative ideological views tend to share more fake news than others (nearly six times as much as moderates or liberals). Their methods included the following salient features. Firstly, they define fake news as "knowingly false or misleading content created largely for the purpose of generating ad revenue" (6), which is a sensible definition given that Facebook and its content creators' primary means of profit is through advertising. Secondly, they concede that "[g]iven the difficulty of establishing a commonly accepted ground-truth standard for what constitutes fake news, our approach was to build on the work of both journalists and academics who worked to document the prevalence of this content over the course of the 2016 election campaign" and "used a list of fake news domains assembled by Craig Silverman of BuzzFeed News, the primary journalist covering the phenomenon as it developed." A website is considered fake news if it has "the hallmark features of a fake news site: lacking a contact page, featuring a high proportion of syndicated content, being relatively new, etc." (6). Given that little explanation is provided as to why we should consider 'hallmark features' or trust these researchers as epistemic guides, one may be concerned that this may foster elitist perspectives about what is fake news in a manner that lacks awareness of potential biases. Thirdly, they employed supervised machine learning techniques to discriminate this sample of fake news websites from those which are merely politically hyper partisan, producing a list of 495 websites of which any article shared by a Facebook user from these sites is considered fake news (7). This algorithm drew from fake news sources curated by the Silverman List of fake news websites, where the classifier sifted through 7 million web pages over 6 months. Here we observe deference to a set of purported epistemic elites to adjudicate informational quality.

To use a second example, Zubiaga et al. (2016) study the dynamics of rumours, defined as unverified and yet instrumentally relevant information in circulation whose veracity is questionable. Several different kinds of rumours are analyzed: 'pipe dream' rumours (wishful thinking), 'bogy' rumours (increase anxiety and fear), and 'wedge-driving' (promote hatred). Using Twitter's application programming interface, the authors employed journalists to manually categorize sets of Twitter threads into 'true' or 'false' stories, and assigned a threshold number of re-Tweets as the minimum number of re-Tweets required to be considered a proper rumour. A 'resolving Tweet' is a Tweet that ultimately establishes the truth-value of the rumour. Tweet topics ranged nine events, including discussions of the 2015 Charlie Hebdo massacre, 2014 Ferguson protests, and the Ebola-Essien hoax, given these events' importance to the general public at the time. Tweets were annotated and categorized by a set of epistemic elites who were chosen to determine informational quality, consisting of three PhD students, one postdoctoral researcher, and members of the public conducting classification tasks on crowdsourcing website CrowdFlower, leading to a total of 233 annotators, most of whom were from the public. What is salient about this study's methods is that what constitutes poor quality information, in the form of rumours, is decided primarily by a combination of a research group's own members and secondarily by a non-randomized sample from the

https://doi.org/10.1017/psa.2023.9 Published online by Cambridge University Press

public through Crowdflower.

As a third example, Murayama et al. (2021) provide a Time-Dependent Hawkes process model of fake news spreading on Twitter via a machine learning classifier trained on Tweets that are fact checked with respect to the websites Politifact.com and Snopes.com between March and May in 2019. Their model posits the following wave equation governing the initial spread of misinformation, which is typically an initial global peak in the distribution, and the subsequent attempts by Twitter users to correct that piece of misinformation leading to a local peak later in time. More precisely, this 'cascade model' of information flow posits that the probability of the news item being shared in an interval $[t, t+\Delta t]$, for some time t, is $\lambda(t)\Delta t$. Here, $\lambda(t) = p(t)h(t)$ and p is a function of t defined as:

$$p(t) = a\left[1 - rsin\left(\frac{2\pi}{T_m}(t+\theta_0)\right)\right]e^{-(t-t_0)/\tau}$$

and

$$h(t) = \sum_{t_i<t} d_i\phi(t–t_i)$$

In this model, $p(t)$ is the oscillation of discussion of a news Tweet (parameterized by $T_m = 24$ hours), a, r are real-numbered constants describing amplitude, θ_0 the phase, τ the time constant of decay, t_i is the time of the $i-$th post, d_i is the number of followers of the $i-$th post, and ϕ is a heavy-tailed memory kernel of the time-lag between the initial fake news post and the later correction item:

$$\phi(s) = \begin{cases} c_o & 0 \leq s \leq s_0 \\ c_0(s/s_0)^{-(1+\gamma)} & s_0 < s \end{cases}$$

for s some time, and empirically discerned constants $c_0 = 6.94 \times 10^{-4}$ seconds, $s_0 = 300$ seconds (5 minutes), and $\gamma = 0.242$. This mathematical model attempts to discern the *dynamical structure* of misinformation propagation through internet discourse. The model's empirical adequacy is claimed to be demonstrated through several datasets of fake news on Twitter, evaluated by the dictates of a set of fact-checking websites, and where key parameters are contingent upon researchers' operationalizations of misinformation. The model's secondary purpose is also to predict future instances of fake news via machine learning classification methods.

Summarizing these studies' methods, a typical epistemic workflow procedure for studying information deprivation consists, firstly, of a set of entities chosen whose informational content we seek to assess (e.g. texts from Tweets). Secondly, a level of analysis is defined in which the information content is given structure so that analysis can proceed appropriately (e.g. as a proposition, as a photo, as a normative statement, etc.). Thirdly, data at that level of analysis is initially collected and categorized by human beings who typically a set of chosen epistemic elites, and only sometimes general members of the public. Fourthly, the data is fed into computer algorithms to either, for

https://doi.org/10.1017/psa.2023.9 Published online by Cambridge University Press

instance, train neural networks to categorize novel data sets, which would hypothetically enhance the predictive accuracy of these algorithms' ability to detect future tokens of categorized types of information, or to conduct statistical analysis.

2 Informational Norms by Agreement

Having surveyed the core methodology of mainstream computer-assisted information deprivation studies, we outline several issues with these methods. Firstly, the concept of truth employed in many studies is frequently either ambiguous or lacking in rigour with respect to its epistemic foundations. For instance, some social scientists such as Vosoughi et al. (2018, 1) explicitly acknowledge how value-laden the term 'fake news' is: "[T]he term has lost all connection to the actual veracity of the information presented." However, they nonetheless claim that fake news and misinformation can be understood using "the more objectively verifiable terms "true" or "false" news" and yet do not specify what these terms mean or refer to. It is typically implicitly taken for granted that (a) value judgments are insufficiently intertwined with alethic (fact-based) judgments to undermine the purported objectivity of such judgments; (b) there are clear facts of the matter as to what the truth is; and (c) that such judgments can be adequately translated into computer code that ensure construct validity is preserved from the initial human-based model of information deprivation to the algorithm's method of discerning novel cases of information deprivation.

Each of (a) - (c) presents significant potential theoretical and practical problems; we focus on both (a) and (b) before returning to (c) later. Information deprivation studies are conducted such that the units of analysis are not only intrinsically socially constructed but are also not the kind of entities that admit of a representation relationship of the kind typically found in models of the natural and other social sciences. For instance, in physics it is common to present a set of differential equations with observable terms having clearly defined measurement procedures discerning observable entities and relations within physical phenomena that are embedded within nature itself. These mathematical models are comparatively value-neutral given that physical measurements are contingent upon intersubjectively verifiable observable outcomes. While there are no doubt background theoretical virtues, such as parsimony, explanatory power, and logical consistency, and moral and aesthetic values, which guide the context of scientific discovery and justification (Douglas 2009), such values are comparatively less present in the physical sciences than they are in the case of information deprivation studies. The situation is more complex in the latter case given that what constitutes information, and what constitutes deprivation, are intrinsically value-laden properties that cannot be directly empirically observed given that information is socially constructed and interpreted. What is observable are tokens of informational types (e.g. words, photos, etc.) and not the value judgments themselves; the choice of what constitutes an appropriate level of analysis for a piece of information is largely contextually defined by a background set of informational preferences which are continually up for societal discussion. Construct validity will critically depend on what various stakeholders will consider in-

https://doi.org/10.1017/psa.2023.9 Published online by Cambridge University Press

formation deprivation; this relationship is largely absent in the physical sciences where value judgements do not typically change the ontology of the phenomena.

Concerning (b), while debate about information quality has centered around the concept of truth, it is unclear whether the concept is needed when we can retain normativity about information quality in other ways. I propose we instead focus on the concept of 'convergence of agreement' to use information in certain ways that are either rewarded with integration into social convention or disincentivized by social sanction. Information deprivation should be understood rather with respect to a set of informational norms that citizens can consent to agree upon as reasonable ways in which to interpret information. That is, other non-alethic values should be emphasized as a criteria of informational quality in a manner which retains sensitivity to citizen values.

To see why agreement rather than truth is the superior concept to use for information deprivation studies, note the following. Firstly, any useful scientific theory will necessarily employ false idealizations to enhance predictive power; the widespread usage of calculus in physics is an example of explicitly false but profoundly useful mathematical models which posit differentiable functions whose domain of application (e.g. the study of particles in physical space) is already known to not be continuous in nature (which is required for calculus to be isomorphic to reality). And yet, it would surely be unreasonable to call physics, and other natural sciences which employ idealizations, a case of information deprivation. Secondly, if one wishes to characterize information in alethic terms, one is confronted with pessimistic meta-induction arguments to the effect that large portions of the history of human knowledge are considered false. However, it is arguably unreasonable to refer to nearly all of our previously held beliefs as information deprivation even if they were false and harmful, given that misinformation is best understood as a relational phenomenon relative to the best epistemic practices of the time period. Hence, it is not reasonable to call most prior beliefs misinformation in this sense. Thirdly, misinformation studies often focus on political issues in which there are genuine disagreements admitting of a plurality of conflicting and yet equally legitimate answers on the same topic. This is not to say that all positions in a debate are uniformly distributed with respect to their justification; rather, purported cases of information deprivation cannot ignore the diversity of background informational values. For instance, it has been argued that many United Nations agencies and non-governmental organizations have provided inaccurate estimates of the safety of refugees, were they to return to their home countries, that are tantamount to misinformation (Gerver 2017). However, whether this constitutes a case of misinformation is entirely dependent upon the level of risk a refugee is willing to take in returning to their home country and what international observers judge to be appropriate risk tolerances as well. There is no fact of the matter as to whether there is misinformation here given a genuine plurality of values and divergence of how to weight these values. Many cases of purported information deprivation have structural similarities to issues like these, especially in political matters.

It is therefore more methodologically prudent to define information deprivation in some other manner than merely with respect to its purported alethic status. This arises

https://doi.org/10.1017/psa.2023.9 Published online by Cambridge University Press

in the context of (c) the construct validity of information deprivation models in machine learning. Returning to the aforementioned model of Murayama et al. (2021), notice how the authors defend the construct validity of this model. In this model, a set of Tweets is initially chosen for analysis by the researchers, given their belief that citizens would find the topics of these Tweets of social relevance and concern; this already illustrates a subjective bias in the epistemic supply chain where researchers already have decided what constitutes important (mis)information. Secondly, observable terms of the model are defined and the wave equation's mathematical structure is hypothesized. Positing a model with an initial cascade followed by a smaller cascade of re-Tweets is a model whose structure has been decided upon by an initially chosen definition of misinformation; hence, alternative operationalizations would lead to alternative mathematical structures that need not have the structure of misinformation of other operationalizations. There is lack of measurement robustness in information sciences given that multiple independent measurement procedures are unlikely to converge on their measurement outcomes concerning information deprivation, especially in dynamical models like this. Thirdly, whether a Tweet is misinformation or not is decided simultaneously by researchers and a specific sub-subset of the general public who both happen to have engaged with the previous fake news item and who are willing to participate in the study and criticize its informational quality. Here, both researchers and the general public must input their values together to come to a judgment that the Tweet's topic is informationally deficient; it follows that the the construct validity of the classifier's ability to classify future cases of informationally deficient Tweets is itself contingent upon a continually updating set of both groups' informational values. Indeed, the exact structure of the wave-equation's parameter values would change depending on what both the public and researchers believe constitutes fake news, thereby intrinsically altering the model's success conditions. This suggests that confidence should be lowered regarding the purported mechanical objectivity of the classifier's algorithm.

Notice that this epistemic supply chain is radically distinct from those in the physical sciences, where values typically do not play as direct a role in adjudicating the empirical adequacy of the model. Hence, the predictive accuracy and explanatory power of this model is intrinsically tied to the continually updating and contingent set of informational preferences of both researchers and average citizens alike. Only once such preferences are solicited and set constant can measurement proceed in a manner which has initial construct validity; as it stands, the methodological emphasis of epistemic elites privileging concepts of objective truth typically ignores other salient features of information deprivation, given that 'objectivity' connotes non-negotiability in a way that is inconsistent with the protean nature of citizen deliberation. Citizens' informational preferences are typically not alethic in nature and often concern whether or not a news item provides enough context, whether it is overly emotionally provocative in an unconstructive manner, whether it makes light of an otherwise serious event, whether it inappropriately frames a perspective bordering on defamation, or whether it speaks to the concerns of all relevant stakeholders (Sunstein 2020). Indeed, sensitivity to the value-ladenness of the epistemic supply chain has led some information deprivation scholars to define

https://doi.org/10.1017/psa.2023.9 Published online by Cambridge University Press

'disinformation' as "misleading information that has the function of misleading," with an emphasis on function (Fallis 2015, 422). Since the function of information is not a property persisting in nature itself, but is a value-laden property of information relative to the epistemic goals of users and their community, it follows that what counts as information deprivation is intrinsically tethered to the needs and preferences of society's members.

3 Democractic Engagement and Informational Norms

To see how this shift away from a methodological emphasis on the concept of truth, and towards agreement, would work in practice, notice the following potential methodological virtues. Firstly, we could enhance the democratic features of information science by asking for input from citizens as to their informational needs and desires, as it is after all citizens' well-being that social scientists are trying to serve. Secondly, information scientists would be able to enhance trust in the general public in social scientific findings pertaining to informational matters if they remained neutral on matters of fact, letting citizens and other scientific fields decide what matters to them concerning informational quality. This is crucial as it is not the business of information scientists to tell us what is true or even necessarily what misinformation is; this responsibility is for the information ethicist. Rather, given a certain citizen consulted conception of the kind of information we want to have disseminated in society, the information scientist ought to (i) study novel forms of information deprivation, as defined by citizens; (ii) discern the structure of information flow and consumption; (iii) provide suggestions for mitigating further information deprivation. The information scientist therefore functions as a form of epistemic hygienist explaining the dynamics of information while remaining comparatively agnostic on what a healthy information polity looks like.

There remains little dialogue between citizens and social scientists about how to measure and operationalize measures of information deprivation, which can undermine the construct validity of the operationalizations given the potential for a form of epistemic authoritarianism in which social scientists impose their conceptions in a top-down fashion. In practice, many online websites purport to do the job of epistemically policing sources but are trusted merely given their ability to convince others that they are evaluating material in a sufficiently neutral and rigorous manner. However, this method is tenuous in that there is little reason to believe that fact checking websites have the requisite set of experts who are either knowledgeable or credentialed enough regarding the tremendously large variety of topics that such websites discuss. In fact, there is preliminary evidence that laypeople ratings of fake news websites are strongly correlated ($r = 0.9$) with professional fact-checkers' judgements, raising serious questions about the extent at which purported fact-checkers have greater expertise than non-experts (Pennycook & Rand 2019). Soliciting citizen feedback on how citizens understand misinformation could enhance the empirical adequacy of measurement constructs of social scientists' models of information deprivation.

One might protest that such a participatory form of democratic engagement is liable

https://doi.org/10.1017/psa.2023.9 Published online by Cambridge University Press

to succumb to other epistemic pitfalls, given that lay citizens are quite ignorant of many important basic facts that only either highly educated people or epistemic elites are in a better position to know about. For instance, Lupia (2016) has argued at length how many US citizens do not understand the vast majority of laws and even how the government and its election cycles function. This would suggest that citizens who, for example, protested the 2020 US election as a case of electoral fraud are victims of straightforward information deprivation and should not be trusted to adjudicate informational quality. Hence, there seems little reason to suppose that citizens would be in a position to know enough to make important judgments about more sophisticated topics requiring high standards of informational integrity, such as the extent at which climate change is a problem or whether COVID-19 is a serious epidemiological threat.

However, governments could not practically be in a sufficient epistemic position to know the preferences of each of its members regarding economic, moral, and informational matters that could justify imposing a top-down model that could order citizen preferences rigorously (Hayek 1944, 60). After all, recent empirical studies in behavioural economics have shown that informational preferences are surprisingly obscure in that citizens are unable to provide clear and coherent reports about their 'willingness to pay' to obtain, or be hidden from, important or sensitive information (Sunstein 2020). Therefore, a hybrid procedure between the epistemic anarchy of the general public and deference towards epistemic experts is needed to ensure a more mature science of information deprivation. Lupia (2016, 54) has suggested that epistemic elites can assist lay people in their epistemic decision-making without unilaterally making decisions for them. For instance, judges could teach jury members of civil trials a set of conceptual tools to interpret evidence that are sufficiently content neutral so as to be unbiased and yet enhance the probability of these jury members (average citizens) arriving at reasonable verdicts. This participatory model of the epistemic supply chain for ensuring the construct validity of information deprivation models will remain challenging to execute in practice, but may nonetheless prove to be methodologically superior to status quo methods.

4 Conclusion

Soliciting more of the public's conceptions of information deprivation will likely enhance the predictive and explanatory power of models of information deprivation by enabling researchers to formulate better theories of informational discourse. Currently, information deprivation studies typically function as models constructed by *de facto* epistemic elites whose failure to sufficiently consult the general public can lead to unhelpful epistemic bigotry and ignores the deeper complexities of what constitutes and causes information deprivation.

References

[1] Alenezi, Mohammed N. Zainab M. Alqenaei. 2021. Machine Learning in Detecting

https://doi.org/10.1017/psa.2023.9 Published online by Cambridge University Press

COVID-19 Misinformation on Twitter. *Future Internet* 13 (244):1-20.

[2] Britz, Johannes. 2004. "To know or not to know: a moral reflection on information poverty." *Journal of Information Science* 30 (3):192-204.

[3] Bryceson, Deborah. Jesper Jønsson. Richard Sherrington. 2010. Miners' magic: artisinal mining, the albino fetish and murder in Tanzania." *Journal of Modern African Studies* 48 (3):353-382.

[4] Daston, Lorraine. Peter Galison. 2007. *Objectivity*. New York: Zone Books.

[5] Douglas, Heather. 2009. *Science, Policy, and the Value-Free Ideal*. Pittsburgh: University of Pittsburg Press.

[6] Fallis, Don. 2015. "What is Disinformation?" *Library Trends* 63 (3):401-426.

[7] Gerver, Mollie. 2017. "Misinformation as Immigration Control." *Res Publica* 23: 495-511.

[8] Guess, Andrew, Jonathan Nagler, Joshua Tucker. 2019 "Less than you think: Prevalence and predictors of fake news dissemination on Facebook." *Science Advances* 5 (5686):1-8.

[9] Hayek, Friedrich. 1944. *The Road to Serfdom*. New York: Routledge.

[10] Lupia, Arthur. 2016. *Uninformed*. New York: Oxford University Press.

[11] Murayama, Taichi, Shoko Wakamiya, Eiji Aramaki, Ryota Kobayashi. 2021. "Modeling the spread of fake news on Twitter." *PLoS ONE* 16 (4):1-16.

[12] Pennycook, Gordon. David G. Rand. 2019. "Fighting misinformation on social media using crowdsourced judgments of news source quality." *PNAS* 114 (7):2521 -2526.

[13] Sunstein, Cass. 2020. *Too Much Information*. Cambridge: The MIT Press.

[14] Vosoughi, Soroush, Deb Ry, Sinan Aral. 2018. "The spread of true and false news online." *Science* 359:1146-1151.

[15] Zubiaga, Arkaitz, Maria Liakata, Rob Proctor. et al. 2016. "Analysing How People Orient to and Spread Rumours in Social Media by Looking at Conversational Threads." *PLOS ONE* 11 (3):1-29.

https://doi.org/10.1017/psa.2023.9 Published online by Cambridge University Press