Michael Polányi's fiduciary program against fake news and deepfake in the digital age

Zsolt Ziegler[1] ￼

Abstract

This paper argues that Michael Polányi's account of how science, as an institution, establishes knowledge can provide a structure for a future institution capable of countering misinformation, or fake news, and deepfakes. I argue that only an institutional approach can adequately take up the challenge against the corresponding institution of fake news. The fact of filtering news and information may be bothering. It is the threat of censorship and free speech limitation. Instead, I propose that we should indicate reliable information with a trademark and news signing-approved information and brand equity. I offer a method of creating a standard for online news that people can rely on (similar to high-quality shopping products).

Keywords Polányi · Fiduciary program · Fake news · Deepfake · Misinformation · Journalism

1 Introduction

This paper argues that Michael Polányi's account of how science, as an institution, establishes knowledge can provide a structure for a future institution capable of countering misinformation, or fake news, and deepfakes. I argue that only an institutional approach can adequately take up the challenge against the corresponding institution of fake news. In a fiduciary program, knowledge is justified in a way that is contingent upon the "personal participation of the knower in the known and on his view that 'personal knowledge' is not merely subjective but is an act of comprehension described as a 'responsible act claiming universal validity'" (Mullins 2001, 79). At this juncture, either we develop a fiduciary program for internet knowledge, and so we allow space for critical thinking, or we do not. I propose that the stakes are such that if we do not, our democratic way of life may be compromised. The former case would necessitate a new system of safeguards is needed, implying a set of values and traditions. The fact of filtering news and information may be bothering. It is the threat of censorship and free speech limitation. Instead, I propose that we should indicate reliable information with a trademark and news signing-approved information and brand equity. I offer a method of creating a standard for online news that people can rely on (similar to high-quality shopping products).

In this paper, I describe state of the art research on fake news and the steps being made to combat the problem. I then describe a dilemma raised in connection with the extended mind thesis, which I will argue provides an even more pressing issue concerning the state of critical thinking in the age of fake news. Finally, I will briefly describe Michael Polányi's fiduciary program in an effort to outline the framework for an institution capable of fighting against fake news and deepfakes.

2 Fake news and deepfakes in the digital age

The rise of misinformation or fake news poses a challenge not only to the scientific community and those tasked with developing filtering software, but also to the free and democratic world. It has been shown that the primary platform of fake news is social media (Weedon et al. 2017). The implications of the prevalence of fake news on social media become even starker when we consider that approximately 47% of Americans reported that their dominant source of news is social media (Shearer and Gottfried 2017). Clearly, an appropriate system of safeguards is needed.

✉ Zsolt Ziegler
zsolt.ziegler@filozofia.bme.hu

1   Department of Philosophy and History of Science, Budapest University of Technology and Economics, Budapest, Hungary

I define fake news, following Gelfert, as "the deliberate presentation of (typically) false or misleading claims as news, where the claims are misleading by design." (Gelfert 2018, 108). Although fake news itself poses a challenging situation, a new phenomenon known as the 'deepfake' has emerged, exacerbating the problem in the digital age. Deepfake videos are manipulated video clips. Image search engines, social media websites, and public video footage are used to insert someone else's face onto preexisting videos (Guera and Delp 2018). Initially, such videos rose to popularity in the form of entertaining software applications that digitally swapped one person's face with that of another person. However, similar technology can pose a threat when it begins to compromise the reliability of public figures (Suwajanakorn et al. 2017). The increasing prevalence of deepfakes could mean that not only can written news be manipulated to mislead public opinion deliberately, but video and audio sources may no longer provide reliable information either. As Hany Farid (2019) said that: "Imagine a world now, where not just the news that you read may or may not be real—that's the world we've been living in for the last 2 years, since the 2016 elections—but where the images and the videos that you see may or may not be real." Until recently, only minimal research has been done on this visual aspect of the fake news detection problem (Shu et al. 2017).

Several software developers have already taken on the challenge of detecting fake news by applying embedding methods, neural networks that capture malicious information. The development of programs to detect fake news seems to be trending upwards in the scientific community. In the last few years, numerous competitions for developing technology to detect fake news have been announced in the field of computer science. However, according to Jwa et al. (2019), these challenges have not resulted in any significant success and can be largely improved.

Shu et al. (2017), on the other hand, seem to be more optimistic after reviewing the state of fake news detection. They claim that fake news detection using composite network analysis combined with current machine-learning techniques can create a new, more generic ways of defining fake news that will enable easier detection by algorithms. Conceptualizing fake news in a way that allows it to be identified in different contexts by algorithmic intelligences "would ease future metamodelling of the entry object and enable better generalistic misinformation detecting agents to be manufactured" (Cardoso Durier da Silva et al. 2019, 2768).

Countering fake news and deepfakes can be done in two ways. On the one hand, self-organizing initiatives may try to take on the challenge of filtering fake news. People with the necessary skills and knowledge can form self-organized interest groups and create independent software to counter fake news. In this case, there is no need to institutionally organize groups, as they share a tacit commitment to fight against manipulation. The fiduciary program that I will outline in Sect. 5. could also be applied to self-organized attempts at combatting fake news. I will describe this program in greater detail shortly, but for now it will suffice to note that, according to a fiduciary program, a shared system of values is needed to establish an institution that can counter fake news and deepfakes. This system must be comprised of certain values rooted in mutual trust, resulting in a joint commitment to truth. An outlook which favors self-organized models contends that this system of values may evolve naturally, establishing a system that can take up the gauntlet against fake news.

On the other hand, one might argue that the organized creation of fake news needs an organized countering system. There may be economic, social, or political benefits to altering public opinion. For instance, Lazer et al. (2018) raised concerns over whether Russia was able to manipulate some of the major social media platforms during the 2016 U.S. election. According to recent congressional testimony (Senate Judiciary Committee 2017), this possibility cannot be excluded. Therefore, it would be naive to think that powerful stakeholders would not be motivated to further advance fake news as well as deepfakes. Cases such as the relative popularity of the Pizzagate conspiracy theory, which made baseless claims about a human trafficking child abuse organization linked to Hilary Clinton's presidential campaign, demonstrate the potential gains to be had for different power interests if they are motivated to create and amplify fake news (Fisher et al. 2017).

I argue that only an institionalized system can counterbalance the destabilizing potential of fake news. Suppose that any stakeholder possessing significant infrastructural and financial resources could create fake news to serve their own interests. Powerful stakeholder would be able to direct their resources in one specific direction with the aim of manipulating public opinion. Self-organized-independent groups may attempt to counterbalance fake news, but without a centralized program with infrastructural and financial support these can hardly exert enough strength against potential organized creators of fake news. Therefore, we cannot count on naturally developing self-organized initiatives to create countermeasures.

Another factor that needs to be mentioned with regard to fake news is a fundamental human characteristic described within social psychology: one's system of attitudes tends to organize incoming information in a way it best fits one's own internal view of how things are. Agents are unconsciously motivated to defend their system of knowledge, and are less effective at incorporating information that does not fit well into their existing schema of beliefs (Fazio 2000; Khan et al. 2013). Hence, any information—be it fake or not—is better processed if it fits into

one's system of attitudes and ideology, according to the selective exposure hypothesis (Frey 1986).

In the digital age, public opinion has become more vulnerable than ever. Facebook estimates that about 60 million bot programs (automated accounts impersonating humans) populate its platform, capable of magnifying the spread of fake news by orders of scale (Shao et al. 2018). In light of such advances, Lazer et al. proclaimed that "a new system of safeguards is needed" (Lazer et al. 2018, 1094).

3 Epistemically unhygienic minds

Deepfakes and fake news proliferate under certain conditions wherein people tend to endorse information uncritically without any conscious reflection. In the case of fake news, one might fail to consider whether a given piece of information is coherent or has a reliable source. It is even worse in the case of deepfakes, when people not only neglect to be critical but are unable to detect manipulation. To the unaided eye, fake video clips are quite often indistinguishable from real ones, and thus do not provoke doubt in the viewer. In this way, agents are automatically inclined to believe information received from deepfakes.

Clark (2008) argues that automatically endorsed, unreflected-upon information gained from external devices involved in the cognitive process can form knowledge. According to the extended mind thesis (Clark 2008; Clark and Chalmers 1998), external objects found in particular environments can function as a part of the mind. It is the view that some content of cognition might extend across minds, bodies, and certain aspects of the physical environment. This thesis makes it possible to hold that if a physical object plays a particular role in aiding cognitive processes, then it generates genuine mental content. I do not wish to examine the extended mind thesis in detail, but rather to point out a possible troublesome consequence: if we accept Clarks's position and consider that the internet is full of unreliable information, then we face an unpleasant result that our entire extended mental activity might be infected by fake news.

This is the problem of epistemic hygiene. According to Clark et al.:

> "(e)ither the agent consciously encounters some new resource as an ongoing object for various forms of epistemically hygienic practice (such as understanding why it is a reliable source of information) or not. If she does, this makes the resource look, at that moment, more like external equipment (it may then be a source of knowledge while failing to be part of HER) if she doesn't, it looks unable (from the perspective of theories of knowledge that subscribe to some form

of 'awareness condition') to act as a source of knowledge." (2018, 334)

The dilemma supposes that if one doubts the reliability of information gained from external memory and thereby tests it, this action alienates her belief-forming process from her. If, on the other hand, she uncritically endorses the information, she fails to satisfy the awareness condition that is necessary to guarantee knowledge instead of mere belief. Palermos (2014, 1942) suggests an intuitive requirement for a healthy mental life: that one must have the awareness to notice circumstances in which their belief process is unreliable. Pritchard (2010, 144) similarly points out that if the agent does not satisfy this awareness condition, her true belief cannot be a basis for forming knowledge.

Contrary to the awareness condition, Clark argues that some of our mental processes are not conscious, e.g., associative learning mechanisms or implicit biases. (Bargh 2006, 208). He notes that, "given the scope and power of unconscious processing, it should come as no surprise to learn that some forms of epistemic hygiene may likewise be nonconsciously acquired and non-consciously deployed" (Clark 2015, 3769). What Clark is insisting upon is that even if the agent is unable to provide positive reasons for her belief, she may be justified in her confidence in the reliability of her thoughts.

This fairly optimistic view imagines that the brain's plasticity may extend to incorporate external devices of cognition. For instance, the use of an abacus triggers an apt precision assignment that selects certain neural circuits and actions that manipulate the beads of an abacus in such a way that "the resulting perception–action cycles solve a mathematical problem" (Clark 2015, 3771). Of course, it is quite difficult to see whether a mechanism like this could ever spot fake information on the web. Clark's point is rather that we, as humans, have created certain institutions through cultural evolution that establish our collective and individual grip on the world. These institutions then encourage practices that develop critical thinking.

4 Exposure of the public

Unfortunately, the presence of fake news and deepfakes does not portend a bright future in which reliable sources of information are guaranteed institutionally. On the contrary, as I have already suggested, once power and politics are motivated to alter public opinion using fake news and deepfakes, no bottom–up initiatives can counterbalance the influx of misinformation. It seems that what is becoming institutionalized is the fake itself (Senate Judiciary Committee 2017). Therefore, we cannot count solely on the natural

development of epistemically hygienic cognitive processes that unconsciously justify truth and knowledge.

People often lack specific profession-related knowledge and can only rely on experts' opinions and guidelines. However, in the digital age, it is not clear which sources of information are reliable and trustworthy Kovács (2020). This compromises the ability of people to form their opinions circumspectly. Established scientific and journalistic institutions now have too many competitors, blurring understandings of what sources can be trusted.

5 Social epistemology

Before turning to Polányi's fiduciary program, we need to situate Polányi's general epistemic approach, known as social epistemology. In contrast with standard epistemology, which focuses on how agents' individual cognitive processes are capable of acquiring knowledge, social epistemology looks at how knowledge can be acquired from others. It does not wish to account for a procedure guaranteeing justified true belief (Danka 2009, 400–401). Rather it challenges the concept of truth, saying that facts are not discovered by certain scientific methods but instead constructed. Rorty, among others, argued for that the idea of objective truth was nonsense, writing that at best what we can have is a notion of the "social justification of belief" (Rorty 1979, 170). According to social epistemology, for our current purposes, it is enough to say that knowledge is a result of a social fabric.

Social epistemology attempts to explain the reliability of claims by soliciting the opinion of other epistemic agents. This testimony-based belief is a process in which one weighs reasons and evidence received from reliable sources. Social epistemology looks at how an epistemic agent can know when they are justified in trusting an assertion made by others. Polányi's answer, as I will explain, is the institutionalized (fiduciary) attitude of researchers seeking the truth.

The phenomenon of fake news is not one singular and isolated cases. It is at least possible, or even present today, that another institutionalized undertaking is set up to maliciously and systematically mislead public opinion. Most of the information we receive is from testimony from sources that we may believe to be reliable, without knowing for sure. In the past, the good old fashioned prestigious sources of news guaranteed reliability.[1] Now, however, the proliferation of online information sources are under no institutionalized

and socially tested control. For now, testimony, the main way in which we understand public affairs, is no longer a reliable justification.

6 Michael Polányi's fiduciary program offering a solution

As I have argued, no bottom–up initiative can counterbalance the information provided by certain powers to mislead public opinion. Some researchers have been working to develop filtering programs, as I have discussed. These attempts, however, are still relatively unorganized, while fake news can follow a designed strategy.

I see only one chance for fighting against fake news and deepfakes. To this end, I am arguing that Polányi's framework of the fiduciary program can be interpreted as an epistemic enterprise. The fiduciary program offers a certain schema for the collaborative work of scientific actors that might lead to the required institution for correctly informing the public. Michael Polányi has developed a comprehensive view of how the scientific institution provides reliable knowledge through personal knowledge and as well as through the scientists' commitment toward seeking the truth.

The term 'fiduciary' expresses the core idea of the program, suggesting that scientific research is based on mutual trust—involving commitment, community, and shared goals—among scientists. According to Polányi, scientific conduct involves one committing oneself to the truth of knowledge. This commitment motivates the entirety of one's behavior, not only in the research laboratory but also in every aspect of one's personal life. A scientist's actions, however, are pervaded by the knowledge that she could be wrong at any time. Polányi calls this the fiduciary act. The program also presumes that one needs to give up the belief that an infallible and objective core of knowledge is ultimately possible. Nonetheless, personal attachment to seeking the truth is a necessary condition for acquiring knowledge. It is almost a moral duty for the scientist to take her research bone-chillingly seriously. A scientist wholeheartedly focuses on her research. It is her deep commitment toward the subject matter and the weight of responsibility she feels for the research's significance that together make truth possible.

One might ask what, if anything, ensures that the scientist's fiduciary act provides no false beliefs—or, even worse, what if an entire system of scientific belief is false? The only reliable belief-forming processes, Polányi holds, have only been ensured through evolutional development up to now. Therefore, the best that we can do is to follow this strategy. The fiduciary program claims that the evolutionary development of human intelligence resulted in our abilities to observe and know the world around us. We have also learned

---

[1] While a full discussion of the changing status of scientific authority is outside the bounds of this paper, it needs to be mentioned that there are certainly pros as well as cons to the erosion of this authority (perhaps giving more voice to people typically excluded from academia or journalism).

how to rely on those abilities, which are conditions for gaining knowledge (Héder and Paksi 2021). Trusting these abilities is a necessary condition for having knowledge. Polányi sees that knowing reality—not only propositional knowledge but also skills and abilities—is a key factor for any species to survive. Polányi provides a detailed analysis in which he argues that different forms of animal life also developed epistemic skills for survival, although only at a low level (Polányi 2005, 354). He contends that humans naturally have certain intellectual capacities for gaining knowledge guaranteed by evolution. Naturally, since evolution has not ended, these capacities are still being developed as environmental changes are still present.

There is an inexplicable component of human knowledge called tacit knowledge in the Polányian framework. The general aim of Polányi's philosophy is to examine the aspects of tacit knowledge. Polányi believes that humans naturally possess certain intellectual capacities making them able to acquire knowledge that does not need to be justifiable. That is to say, Polányi contends that evolution itself guarantees that at least some of our belief-forming processes are epistemically hygienic. However, these processes are not restricted only to biological organs. Instead, as Polányi emphasizes, scientific research tacitly shares these characteristics of (hygienic) reliable knowledge.

Polányi assumes that not only is our biological make-up subject to the evolutionary development of gaining knowledge, but the entire social and economic structure of the society has been developed in such a way that one is able to access truth. The scientific program, moreover, is capable of differentiating pseudoscience from real science. The structure of the society, with all its institutions and services, bears its current shape in order for us to know and survive. Thousands of years of social evolution have resulted in a contemporary society with institutions and services that guarantee knowledge. According to Polányi:

"Our tacit powers decide our adherence to a particular culture and sustain our intellectual, artistic, civic, and religious deployment within its framework. The articulate life of man's mind is his specific contribution to the universe; by the convention of symbolic forms man has given birth and lasting existence to thought." (Polányi, 2005, 278-9)

And further:

"I must admit now that I did not start the present reconsideration of my beliefs with a clean slate of unbelief. Far from it. I started as a person intellectually fashioned by a particular idiom, acquired through my affiliation to a civilization that prevailed in the places where I had grown up at this particular period of history. This has been the matrix of all my intellectual efforts. Within it I was to find my problem and seek the terms for its solution." (Polányi, 2005, 265)

According to the fiduciary program, personal knowledge is possible only if the scientist recognizes and accepts her physical, social, and institutional embeddedness. Polányi denies that fully objective knowledge is possible. Since tacit knowledge is inherently part of the knowledge, it is simply not possible to separate tacit elements of knowledge from the objective ones. As Polányi puts it:

"It is the act of commitment in its full structure that saves personal knowledge from being merely subjective. Intellectual commitment is a responsible decision, in submission to the compelling claims of what in good conscience I conceive to be true. It is an act of hope, striving to fulfill an obligation within a personal situation for which I am not responsible and which therefore determines my calling. This hope and this obligation are expressed in the universal intent of personal knowledge." (Polányi, 2005, 67)

No one is born with all the socially encoded intellectual capacities which allow science to be conducted. It is one's culture—scientific and social—that shapes one's cognitive character. However, one's questioning of other scientific achievements relies on a scientific method which is saturated with thousands of tacit biases.

Any scientist, having gone through scientific training, learns not only particular objectively explicable parts of scientific knowledge but also takes up the tradition of the scientific institution. This tradition itself is a result of the historical progress by which science has developed its institutions, functions, and practices. All these jointly ensure that one who takes up the scientific fiduciary program will be able to gain knowledge. A shared system of values has been developed in science. According to these values, scientists are collectively accountable for the formation of a scientific community. As the scientific community, the purpose of the entire institution is to compel those values of science that guarantee reliability. These institutional values must, therefore, be tacitly incorporated into one's thinking.

Moreover, the community itself shall implore scientists to follow these virtues. Thus, membership in a scientific community plays a functional role in ensuring scientific knowledge. Science, therefore, is not an extremely liberal institution. Quite the opposite, science determines those practices and ways of thinking that seem appropriate. The evolutionarily developed structure of scientific values ensures that science provides knowledge and differentiates it from pseudoscience.

Science is an exemplar for any institution dealing with information and knowledge. Polányi mentions other institutions that share values systems and accountability
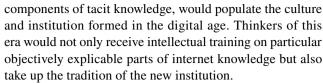
characteristics, for example, the legal and educational systems. Journalism, however, must fall into this category as well because it provides the medium for knowledge to be spread. Journalism used to have the required system of values before the internet era of free information. Similarly to science, there were values that ensured, for example, the reliability of an article because a journalist had to take responsibility for it. In many ways, science and journalism as institutional formations share structural similarities. In fact, they are inseparable components of a whole.

From a social evolutionary point of view, the development of the internet created an unexpected situation that could not be handled by conservatively developed journalism. Since evolutionary progress needs continuous changes to adapt to new phenomena, any groundbreaking revolution brings about a state of affairs that makes it hard for old institutions to continue functioning. Accordingly, fake news created an unexpected situation for established journalism, jamming its well-tried institutions and values. Journalism's lack of experience with malicious information created the opportunity for fake news to flourish.

7 A new system of safeguards

Now, what could an appropriate system of safeguards be, if such a system is possible at all? The answer lies in a structure described as the fiduciary program. More than an organized system of scientific cooperation is needed. Concerning truth and reliability, what is needed is a collective commitment and shift in attitude by researchers, journalists, and content creators. A schema of values, outlined according to the fiduciary program, is needed to establish an institution countering fake news and deepfakes. This system of values may rely on the scientific community's values, but it needs to be extended for journalism. Similarly to science, the new community of internet knowledge—researchers, journalists, and the public—must impose a system of criteria against which any actor may be held accountable. Just as the scientific community upholds the scientific institution along with its tradition, rules, and duties, a fiduciary program of the internet knowledge must create its rules and duties to be imposed on actors within the fiduciary internet program. Online journalism should differentiate true news from fake news in the same way that science differentiates true science from pseudoscience. The new community must pursue a common goal of a well-informed public capable of critical thinking.

Actors within the fiduciary program of the internet knowledge must commit themselves to the truth of knowledge. Personal attachment to seeking the truth would be the central value. Actors shall wholeheartedly focus on their role in establishing the program. Science, several non-explicable

components of tacit knowledge, would populate the culture and institution formed in the digital age. Thinkers of this era would not only receive intellectual training on particular objectively explicable parts of internet knowledge but also take up the tradition of the new institution.

The proposal I have just provided may lack practical steps to build the fiduciary program of internet knowledge. For the moment, however, I do not wish to define precisely those conditions under which the program is feasible. Even in his book, *Personal Knowledge*, Polányi only touches on various elements of science and the social sciences. The program I propose needs to result from a continuous socially developed process. Determining strict conditions that govern the program would only lead to a dead end. The program must be flexible and able to alter its direction according to situational changes.

Nonetheless, what determines the program's success is the attitude of the community—similarly to the attitude of the scientific community. Actors within a fiduciary program of internet knowledge shall share a culture, be part of an institution, and be committed to several moral values toward building a cognitively critical society. Polányi's view is built on three main columns: values, tradition, and commitment. The proposal I have made also lies on these rocks of knowledge.

8 Possible criticisms

One[2] might object that some research suggests that the actual effects of fake news on political attitudes and behaviors are quite minimal. For instance, a study concerning the German election in 2017 indicates that.

> "people's political predispositions strongly orient candidate evaluations and voting preferences, but these predispositions are currently unlikely to be meaningfully altered by attention to DFN (digital fake news). Instead, the effects of consuming DFN through either social media or dedicated fraudulent news sites seem likelier to be that of just confirming and reinforcing people's pre-existing views and biases. (…) the electoral impact of DFN is for now probably quite narrow in scope. (…) we also postulate that selective attention to DFN is possibly a symptom rather than a cause of ongoing ideological polarization." (Leyva and Beckett, 2020, 977)

According to this view, the effect of fake news may not seem that serious. This opinion holds that, though there may

---

[2] I am very thankful to an anonymous referee for highlighting this research on the effects of fake news.

be valid concerns regarding the existence of fake news, the public's opinion cannot be manipulated via fake news. At least for now, "fake news does not crowd out hard news consumption" (Staff 2019; Guess et al. 2021, 24). On the other hand, Zimmermann and Kohring also express more foreboding concerns about fake news in the following:

> "In light of our findings, *one cannot ignore that the success of disinformation is also a defeat for democratic institutions*. While there is certainly not a general loss of institutional trust in Germany, a specific portion of the *German population has become strongly skeptical about legacy news media and the political system over the last years.* From their point of view, *professional journalists and politicians have discredited themselves in covering and dealing with important political topics such as the refugee situation.* As a consequence, these doubly *mistrustful people are yearning for alternative facts for the purpose of orientation and confirmation, with the striking result that the less one trusts in news media and politics*, the more one believes in online disinformation. We thus provided empirical evidence for Bennett and Livingston's (2018) *notion of a disinformation order emerging from a breakdown of institutional trust* and forming in opposition to the established information system. " (italics added) (Zimmermann and Kohring, 2020, 231)

Allcott and Gentzkow likewise showed that "in the aftermath of the 2016 US presidential election, it was alleged that fake news might have been pivotal in the election of President Trump" (Allcott and Gentzkow 2017, 232) It seems that the current state of thought regarding fake news is twofold. On the one hand, many doubt that fake news has significantly compromised democratic thinking thus far, but on the other, there are worrying signs that may warn of the possible threat that fake news poses. The consumption of fake news seemingly applies to "(h)eavy social media users [who] would be more likely to see obscure news sources regularly pop up in their feeds" (Nelson and Taneja 2018, 3723). This is of concern as the amount of time many people spend on social media is certainly increasing (Global Digital 2019 Reports Show Social Media Usage Continues to Rise | YouGov 2021).

Even if we were to grant that fake news has not yet affected public life severely, it is worth drawing attention to the potential problem. The fact that fake news has already been able to affect the opinion of the public should drive us to create countermeasures. I believe a Polányian fiduciary program for internet knowledge is one of the best ways to take up this important work.

Objections may also be raised as to who would constitute such a panel, and whether it would gain widespread acceptance. The fact that many people are skeptical of scientific findings and generally distrust those perceived to be 'elites' creates additional difficulties for the setting up of a fiduciary evaluation.[3]

I believe this issue can be addressed in two ways; however, in their analysis, Mann and Schleifer (2020) show that opposition to science reflects hostility toward scientific elites rather than the scientific method. It means that it is not science per se or scientific knowledge in general that is questioned. Specific scientific issues may become politicized (Dunlap and Jacques 2013) and this results in a degree of skepticism towards those scientific questions, for example, climate change. nonetheless, the moral of scientific method—that is to pursue truth—seems to be intact.

The fiduciary program I offered for internet knowledge relies greatly on the scientific community's joint commitment and attitude towards seeking the truth. Science provides a shared system of values that hold scientists accountable to the entire scientific community, so too shall the fiduciary program do for internet knowledge. Though it is true that some, even in our modern time, are still skeptical of science, hard-core science skeptics represent only a small portion of society (O'Brien and Noy 2020). Various scientific findings have been questioned and have only attained moderate popularity, and on a limited scale. For example, only 2% of the American population believe that the earth is actually flat (YouGov | What the World Thinks 2021). Scientific institution still seems to shape the way the public thinks of scientific matters. If our fiduciary programmed science can do it, then an internet knowledge of fiduciary program might do it as well.

The program I suggest aims to target members of the journalistic community and not the public in general. Polányi's fiduciary program describes the scientific method, which provides accountable knowledge. This program also aims to prescribe a system of values that invoke responsibilities by personally attaching one to the subject they are writing about. Once the majority of content creators commit themselves to the fiduciary program of internet knowledge, the majority of the accessible information will constitute reliable knowledge for the public.

Note further that, from a social evolutionary (Frank 1998) point of view, social practices that play a role contrary to maintaining the work of society will die out with time. If this is true, then the fiduciary program for internet knowledge would ensure that fake news either dies out or gains a positive role.

The scientific institution has developed its methodology and values establishing a fiduciary program ever since the philosophy of nature began. So journalism must be

---

[3] I thank the anonymous reviewer whose comments called my attention to this point.

developing its own methodology for its fiduciary program. The practical aspect for that, however, shall be developed in a decentralized way.

9 Leaving falsehood but approving truth

Some might object that social media companies generate significant profits from news, ads, micro-targeting, and the creation of controversies, bringing groups together even if it be for unwanted consequences. These firms, then, might be unmotivated to filter news. Even when we set up a fiduciary program of internet knowledge that fosters journalism to seek the truth on a grander scale, we may not be able to counterbalance tech giants' interest in click-bait and possibly misinformation. I, however, must admit that the fact of filtering news and information may trigger some negative connotations; a chance for censorship and the limitation of free speech might become a real possibility.

For these reasons, I do not think fake news and certain controversies and disputes shall be banned on the Internet, but a fiduciary program of journalism shall mark the reliable information and news with a branded trademark showing that the information provided is approved. Note that, it is much easier to positively evaluate reliable news than to filter out all false news. With careful planning of a branding strategy, we aim to reach the fiduciary program of internet knowledge's positive brand equity. Brand equity is when consumers react more favorably to the brand than to a generic or unbranded version of the same product (Kotler et al. trademark. 2020, 251–9). In this way, internet users can realize reliable information while unbranded information is still not banned or filtered.

This proposal has a twofold advantage. On the one hand, it does not interfere with tech giants' interests, allowing the presence of various controversies and online disputes on different matters. Those who want to argue for and against conspiracy theories and other exotic issues are free to do it. Click-baiting can continue its journey. On the other hand, it will create a standard (similarly to high-quality shopping products) for online journalism that people can rely on. Over time, people will learn which is the reliable brand. The challenge is to perform the branding strategy of the fiduciary program of journalism. Fortunately, marketers do possess the skill and knowledge to do that.

10 Possible consequences

I realize the program I offer may seem optimistically utopian. A program established on mutual trust and commitment seeking the truth may seem naive. Regardless, I see no other chance for countering fake news and deepfakes.

Only an institutional undertaking can take up the challenge against another institution. I think we face a dilemma: either we develop a fiduciary program of internet knowledge, and so we enable critical thinking, or we do not. In the latter case, our democratic way of life will be compromised. To save it, there must be a new system of safeguards comprised of a set of shared values and traditions that create accountability. The fiduciary community of the internet would create a standard for reliability. However, I must assume that not all actors would join the program, and there would be individuals who will instead create fake news and deepfakes. We must establish a culture that counterbalances these individuals, in both number and competence, via a moral standard. All the society shall be part of the program, making a fiduciary social contract between science, journalism, and the public for a democratic future.

References

Allcott H, Gentzkow M (2017) Social media and fake news in the 2016 election. J Econ Perspect 31(2):211–236. https://doi.org/10.1257/jep.31.2.211

Bargh JA (2006) What have we been priming all these years? On the development, mechanisms, and ecology of nonconscious social behavior. Eur J Soc Psychol 36(2):147–168. https://doi.org/10.1002/ejsp.336

Bennett WL, Livingston S (2018) The disinformation order: disruptive communication and the decline of democratic institutions. Eur J Commun 33(2):122–139. https://doi.org/10.1177/0267323118760317

Cardoso Durier da Silva F, Vieira R, Garcia AC (2019) Can machines learn to detect fake news? A survey focused on social media. https://doi.org/10.24251/HICSS.2019.332

Carter JA, Clark A, Palermos SO (2018) New humans? Ethics, trust, and the extended mind. In: Carter JA, Clark A, Kallestrup J, Palermos SO, Pritchard D (Eds) Extended epistemology. Oxford University Press, pp. 331–351. http://eprints.gla.ac.uk/162579/

Clark A (2008) Supersizing the mind: embodiment, action, and cognitive extension. Oxford University Press, Oxford

Clark A (2015) What 'Extended Me' knows. Synthese 192(11):3757–3775

Clark A, Chalmers D (1998) The extended mind. Analysis 58(1):7–19

Danka I (2009) Practical knowledge versus knowledge as practice. Hum Aff 19(4):397–407

Dunlap RE, Jacques PJ (2013) Climate change denial books and conservative think tanks: exploring the connection. Am Behav Sci 57(6):699–731. https://doi.org/10.1177/0002764213477096

Fazio RH (2000) Accessible attitudes as tools for object appraisal: their costs and benefits. Why we evaluate: functions of attitudes. Lawrence Erlbaum Associates Publishers, Mahwah, pp 1–36

Fisher M, Cox JW, Hermann P (2017) "Pizzagate: From Rumor, to Hashtag, to Gunfire in D.C." Washington Post. https://www.washingtonpost.com/local/pizzagate-from-rumor-to-hashtag-to-gunfire-in-dc/2016/12/06/4c7def50-bbd4-11e6-94ac-3d324840106c_story.html. Accessed 6 Dec 2016

Frank SA (1998) Foundations of social evolution. Princeton University Press, Princeton

Frey D (1986) Recent research on selective exposure to information. In: Berkowitz L (ed) Advances in experimental social psychology, vol 19. Elsevier, Amsterdam, pp 41–80. https://doi.org/10.1016/S0065-2601(08)60212-9

Gelfert A (2018) Fake news: a definition. Informal Logic 38(1):84–117. https://doi.org/10.22329/il.v38i1.5068

Global Digital 2019 reports show social media usage continues to rise | YouGov (2021). https://yougov.co.uk/topics/resources/articles-reports/2019/01/30/global-digital-2019-reports-show-social-media-usag. Accessed 9 Jan 2021

Guera D, Delp E (2018) Deepfake Video Detection Using Recurrent Neural Networks, pp 1–6. https://doi.org/10.1109/AVSS.2018.8639163

Guess A, Nyhan B, Reifler J, Reifler J (2021) Selective exposure to misinformation: evidence from the consumption of fake news during the 2016 U.S. presidential campaign

Héder M, Paksi D (forthcoming) Guide to Personal Knowledge

Jwa H, Oh D, Park K, Kang JM, Lim H (2019) exBAKE: automatic fake news detection model based on bidirectional encoder representations from transformers (BERT). Appl Sci 9(19):4062. https://doi.org/10.3390/app9194062

Khan R, Misra K, Singh V (2013) Ideology and brand consumption. Psychol Sci 24(3):326–333. https://doi.org/10.1177/0956797612457379

Kotler P, Armstrong G, Harris LC, He PH (2019) Principles of marketing, 8th edn. Pearson Education

Kovács L (2020) Társadalmi marketing és világjárvány. In: Kovács L (ed) Globális kihívás – lokális válaszok. A koronavírus (Covid19) gazdasági és társadalmi összefüggései és hatásai. Savaria University Press, Szombathely, pp 195–206

Lazer DMJ, Baum MA, Benkler Y, Berinsky AJ, Greenhill KM, Menczer F et al (2018) The science of fake news. Science 359(6380):1094–1096. https://doi.org/10.1126/science.aao2998

Leyva R, Beckett C (2020) Testing and unpacking the effects of digital fake news: on presidential candidate evaluations and voter support. AI & Soc 35(4):969–980

Mann M, Schleifer C (2020) Love the science, hate the scientists: conservative identity protects belief in science and undermines trust in scientists. Soc Forces 99(1):305–332. https://doi.org/10.1093/sf/soz156

Meeting | Hearings | United States Senate Committee on the Judiciary (2017). https://www.judiciary.senate.gov/meetings/extremist-content-and-russian-disinformation-online-working-with-tech-to-find-solutions. Accessed 20 Dec 2019

Mullins P (2001) The "post-critical" symbol and the "post-critical" elements of Polanyi's thought. Polanyiana 10:77–90

Nelson JL, Taneja H (2018) The small, disloyal fake news audience: the role of audience availability in fake news consumption. New Media Soc 20(10):3720–3737. https://doi.org/10.1177/1461444818758715

O'Brien TL, Noy S (2020) Political identity and confidence in science and religion in the United States. Sociol Relig 81(4):439–461. https://doi.org/10.1093/socrel/sraa024

Palermos SO (2014) Knowledge and cognitive integration. Synthese 191(8):1931–1951

Polanyi M (2005) Personal knowledge towards a post-critical philosophy. Routledge, London

Pritchard D (2010) Cognitive ability and the extended cognition thesis. Synthese 175(S1):133–151

Rorty R (1979) Philosophy and the mirror of nature. Princeton University Press, Princeton

Shao C, Ciampaglia GL, Varol O, Yang K-C, Flammini A, Menczer F (2018) The spread of low-credibility content by social bots. Nat Commun 9(1):1–9. https://doi.org/10.1038/s41467-018-06930-7

Shearer E, Gottfried J (2017) News use across social media platforms 2017, from Pew Research Center's Journalism Project website: https://www.journalism.org/2017/09/07/news-use-across-social-media-platforms-2017/. Accessed 20 Dec 2019

Shu K, Sliva A, Wang S, Tang J, Liu H (2017) Fake news detection on social media: a data mining perspective. Sigkdd Explorations. https://doi.org/10.1145/3137597.3137600

Staff SL (2019) UC Berkeley researchers develop technique for detecting video simulations. The Daily Californian. https://www.dailycal.org/2019/07/05/uc-berkeley-researchers-develop-technique-for-detecting-ai-video-simulations/

Smith K (2012) Homeopathy is unscientific and unethical. Bioethics 26(9):508–512. https://doi.org/10.1111/j.1467-8519.2011.01956.x

Suwajanakorn S, Seitz SM, Kemelmacher-Shlizerman I (2017) Synthesizing Obama: learning lip sync from audio. ACM Trans Graph 36(4):95:1-95:13. https://doi.org/10.1145/3072959.3073640

The Daily Cal Covers Hany Farid's Deepfake Detection Approach (2019). From UC Berkeley School of Information website: https://www.ischool.berkeley.edu/news/2019/daily-cal-covers-hany-farids-deepfake-detection-approach. Accessed 20 Dec 2019

Weedon J, Nuland W, Stamos A (2017) Information operations and Facebook

YouGov | What the world thinks (n.d) (2021). YouGov: What the World Thinks. Retrieved 24 Apr 2021, from https://today.yougov.com/opi/live_survey_results/6147e107-0a97-11e8-b216-d9ff14e72543/question/cecf46a6-0a97-11e8-bd7b-c532039a14a0/toplines/

Zimmermann F, Kohring M (2020) Mistrust, disinforming news, and vote choice: a panel survey on the origins and consequences of believing disinformation in the 2017 German Parliamentary Election. Polit Commun 37(2):215–237. https://doi.org/10.1080/10584609.2019.1686095