

# **The epistemic status of reproducibility in political fact-checking**

**Alejandro Fernández-Roldán &  
David Teira**

Fact-checking agencies assess and score the truthfulness of politicians' claims to foster their electoral accountability. Fact-checking is sometimes presented as a quasi-scientific activity, based on reproducible verification protocols that would guarantee an unbiased assessment. We will study these verification protocols and discuss under which conditions fact-checking could achieve effective reproducibility. Through an analysis of the methodological norms in verification protocols, we will argue that achieving reproducible fact-checking may not help much in rendering politicians accountable. Political fact-checkers do not deliver either reproducibility or accountability today, and there are reasons to think that traditional quality journalism may serve liberal democracies better.

Fernández-Roldan, A., Teira, D. The epistemic status of reproducibility in political fact-checking. *Euro Jnl Phil Sci* 14, 12 (2024).

<https://doi.org/10.1007/s13194-024-00575-8>

## **1. The rise of political fact-checking**

Political fact-checking (PFC) is a deceptively simple activity: fact-checkers assess the truthfulness of political claims with a formal score, releasing the assessment to the public through internet sites and social media (Amazeen et al., 2018). Like journalists, in the Anglo-American tradition, fact-checkers often present their mission as holding politicians accountable for their claims (Graves & Cherubini, 2016). The implicit rationale is as follows. In complex representative democracies, voters often cannot discern whether politicians are really protecting citizens' interests since there are plenty of opportunities for politicians to disguise their true agendas. Journalists would ideally fill this gap by providing voters with the necessary information about what politicians are doing. There is indeed some preliminary evidence suggesting that quality journalism is important for democracies: for instance, press freedom is often associated with a country's position in international corruption indicators (Breen & Gillanders, 2020). PFC would deliver a simplified, but more effective version of these accountability checks: signalling voters with a score whether a politician is lying or telling the truth in each statement they utter so that voters can punish liars, if they wish.

Fact-checking is almost as old as the world wide web and, already in 1994, we find the general hoax debunking site *Snopes*. PFC is now a global movement originating in 2003 with the launch of Factcheck.org. In 2007, two newsroom fact-checkers were born: *Politifact* (affiliated to the *Tampa Bay Times*) and *Factchecker* (affiliated with the *Washington Post*). The rise of fact-checking sites accelerates in the mid-2010s: the count rose to 113 in 2016 (Graves, 2016), 149 in 2018 (Stencel & Griffin, 2018) , 290 in 2020 (Stencel & Luther, 2020), and 378 active fact-checking sites in 2022 (Stencel et al., 2022)<sup>1</sup>.

There are different accounts about the proliferation of these organizations. A usual explainer is the growth of political *fake news* online (Vargo et al., 2018). Obvious examples are the Cambridge Analytica scandals or the voter microtargeting schemes implemented in both the 2016 American presidential elections and the Brexit referendum (Vicario et al., 2016). But there might have been also older factors at work in the rise of PFC, such as a decades-long disenchantment with traditional approaches to political reporting in the US (Birks, 2019). The 2009 Pulitzer Prize for national reporting awarded to *Politifact*, little after its launch, may have been a signal that it was a path worth exploring.

More than a decade afterward, it is perhaps time to take stock and discuss how much further PFC can go. We will take as our guiding thread an ongoing discussion among fact-checkers: is it just another form of traditional journalistic verification or is it rather a proto-scientific method? Quality journalism has always fact-checked news before publication, but without any standardized procedure<sup>2</sup>. According to some leading practitioners, PFC should be based instead on a verification method that, if correctly implemented, will lead to *reproducible truth scores*: reproducing the verification process should yield, time and again, identical outcomes. This is a degree of impartiality in verification (regarding the personal interests or preferences of the fact-checkers) that standard journalism can rarely achieve. This method would make PFC agencies more trustworthy than conventional journalists since the independence of the latter as

---

<sup>1</sup> Mark Stencel and Joel Luther are in charge of updating the Duke Reporters' Lab database of fact-checking sites every year according to certain criteria that can be accessed at <https://reporterslab.org/how-we-identify-fact-checkers/> (Last access on July 21, 2023)

<sup>2</sup> Throughout this paper we will use *Quality journalism* as an informal shortcut to refer to reporting characterized by features such as: its trustworthiness; diversity in sources, coverage, etc.; depth and breadth of information; comprehensiveness; or its emphasis on public affairs. See (Lacy & Rosenstiel, 2015).

watchdogs of Western democracies is increasingly tainted by commercial interests (Braun & Eklund, 2019). Reproducible PFC would foster politicians' accountability, signalling their lies in a way that would push citizens to punish liars with their votes.

We want to examine here whether reproducible PFC may live up to these expectations: do we have reasons to expect reproducible fact-checking to be better than traditional journalistic verification at holding politicians accountable?<sup>3</sup> As we will see in sections 2 and 3, academic PFC is, indeed, often defended as a proto-scientific endeavour based on reproducible verification protocols. We will present how these protocols should work according to the guidelines of the International Fact-Checking Network (IFCN), an international audit body for fact-checking agencies. In section 4 we will discuss under which conditions reproducible verification protocols may achieve a quasi-scientific status. Drawing on Hasok Chang's account of the history of thermometry, we will argue that reproducibility may indeed pave the way for articulating PFC as a proto-scientific discipline. But, as we will see in section 5, there is a clear tension between the two founding principles of academic PFC, achieving reproducibility, on the one hand, and effective lie detection, on the other.

The verification protocols now sanctioned by the IFCN are based on *standards*, general guidelines open to discretionary interpretation and, therefore, not easily reproducible. Perfectly reproducible verification should be articulated on strict *rules*, such as those in use for automatic fact-checking algorithms. We are going to argue that rule-based fact-checking may achieve proto-scientific objectivity, but it will rarely make politicians more accountable. Standard-based fact-checking is perhaps more suitable for tracking political lies, but it will rarely be reproducible. As of today, it does not seem very likely that PFC, rule- or standard-based, will foster political accountability better than good old quality journalism. As we will argue in our conclusions, we actually have reasons to think that PFC may be doing more harm than good: reducing the traditional verification process to a simplistic score, without clear construction and interpretation rules, will foster, if anything, mistrust in journalism (see: Carson et al., (2022) for an example on how an independent fact-check can backfire and reduce trust in journalism).

---

<sup>3</sup> This paper is the theoretical companion of two empirical surveys trying to grasp the actual effects of PFC on political accountability. The first one is already published: (Fernández-Roldán et al., 2023).

## 2. Political fact-checking: journalism or science?

Following (Hansson, 2018), we may characterize journalism and science as *fact-finding practices*. Practitioners in both fields presuppose a common world, in which there are empirical facts separable from the practitioners' values. These facts may be empirically established through an open communal pursuit, in which an egalitarian uptake of criticism is key. Science, of course, is significantly more systematic and specialized in its endeavours than journalism, but journalists often notice the continuity between both types of practices, sometimes arguing that the adoption of a more scientific stance would improve the quality of their reporting -see (Elías, 2019) for a review and discussion. Not every journalist agrees, of course. The debate around PFC we are going to examine here illustrates this tension. Fact-checkers are split in their self-perception of their discipline: according to Graves & Cherubini (2016), 73% of fact-checking professionals agree either strongly or very strongly on being considered journalists, and only 30% agree with equal intensity to being considered *academics*.

The main argument among PFC for considering their job a standard form of journalism hinges on verification. Early in the 20<sup>th</sup> century, objectivity became a standard for professional journalists (Meyers, 2020): news agencies aimed at readers distributed across the ideological spectrum and adopted a set of style conventions that identified objective reporting, with a crucial emphasis on factual accuracy<sup>4</sup>. *Ex ante* verification (double-checking each source) became an editorial procedure to secure the accuracy of every piece of information published. PFC would transform this *ex ante* process into an *ex post* standalone activity: the claims targeted are now public and the verification yields a formal score (Mantzaris, 2018).

As a matter of fact, the institutional organization of traditional newsrooms and fact-checking agencies is often similar. Fact-checking organisations in countries with high levels of democracy (e.g. Western Europe) tend to be independent companies following the newsroom model, funded by subscription and/or advertising revenue (Graves & Cherubini, 2016). Moreover, the majority of fact-checkers are associated with media organisations (Stalph, 2018) and, despite the huge diversity of the movement, it is

---

<sup>4</sup>(Galison, 2015) explores some analogies in the evolution of the ideal of objectivity in science and journalism. Crucial for our argument below is that public trust seems not to depend anymore on the value-free ideal, neither in science nor in journalism (Elliott, 2017). Still, it is open to discussion how trust in these two institutions should be reconstructed: for a preliminary exploration in line with our own view, see (de Melo-Martín & Intemann, 2018).

reasonable to assert that fact-checking “as both an international movement and a field of practice, is at core journalistic”(Graves & Cherubini, 2016).

Those who consider PFC an academic activity argue instead that some fact-checking agencies have indeed emerged from joint ventures with academia and university-linked foundations (Graves & Konieczna, 2015). E.g., the Annenberg Public Policy Center at the University of Pennsylvania was essential in the birth of the first PFC organisation, Factcheck.org, back in 2003 (Marietta et al., 2015). These university-based organizations have access to funding sources not available to journalists (academic grants). And their approach to verification adopts a social scientific stance: they formally combine scholarly and journalistic expertise or methods (Graves, 2018). By dint of their method and organization, fact-checkers become *experts* more than *journalists* (Graves & Cherubini, 2016).

According to its academic practitioners, PFC certainly goes beyond traditional journalistic verification in one key point: it follows a standardized procedure more reminiscent of a scientific protocol than a journalistic practice. In traditional political reporting there were no precise guidelines to achieve neutrality (Gans, 2004). Quality fact-checking agencies adopt institutional warrants of neutrality and follow a formalized protocol for selecting and scoring claims. We find a stylized illustration of this approach in the Code of Principles of the International Fact-Checking Network (IFCN), a consortium of agencies auditing whether its potential signatory members comply with their agreed best practices<sup>5</sup>.

According to the Code, fact-checking agencies should be legally registered organizations, with a public archive of their work. Their fact-checks should exhibit consistency, leaving no room for partisanship or unfairness (i.e., double standards). There should be conflict of interest policies, explicitly targeting political biases in verification. The sources and the verification method should be public. There should be a correction procedure open to the public. And the funding and the organization of the agency should be transparent.

In short: for some political fact-checkers, verification is just a journalistic fact-finding practice, both in content and institutional organization; for some others PFC is a more academic endeavour. We will not take sides a priori between any of these two views of

---

<sup>5</sup> The IFCN Code of Principles is accessible on: <https://ifcncodeofprinciples.poynter.org/know-more/the-commitments-of-the-code-of-principles> (Last visited on July 21, 2023)

PFC. On the one hand, unlike, say, scientific and pseudo-scientific medicine, the difference between journalistic and academic PFC is a matter of degree. Unlike quacks with patients, journalists have been traditionally effective at holding politicians accountable for their claims. What academic fact-checkers claim is that they can do it better with reproducible verification protocols, and this claim will be the target of our analysis. On the other hand, like most philosophers of science, we do not think that the institutional organization of PFC or its reproducibility criteria are enough to set apart academic PFC as a scientific endeavour -we will return to this point below, in section 3. But adopting scientific manners (verification protocols, truth scales, etc) has certainly an impact on the audiences, at least in those countries where science is perceived as a trustworthy institution and journalists are comparatively less credible. As we will see in our last section, we should wonder whether academic PFC deserves any of this trust.

*Prima facie*, the difference between the journalist and the academic stance is far from trivial. Depending on the self-perception of PFC, a bias in their verification activity would have different consequences. Consider, for instance, a fact-checker focusing more often on right-wing than on left-wing politicians. This is a feature, not a bug of conventional journalism, the editorial line of any news organization signals to its audience the sort of values guiding its journalists. For PFC agencies with scientific aspirations, this attentional imbalance between political parties would instead count as a potential bias, perhaps originating in the ideology of the fact-checker selecting the verifiable claims. The IFCN's model of PFC tries to control for such a bias, as it would happen in any *bona fide* scientific discipline. As we have argued elsewhere, controlling for bias is a pre-requisite for gaining the audience's trust (Teira, 2016). We should then discuss how a verification protocol could do it.

### **3. Reproducible verification protocols**

PFC come in all sorts of flavours: there are many different institutional arrangements and a lot of diversity in their verification practices. As of today, there is no consensus on the superiority of any single approach to PFC (Walter et al., 2020). The closest we come to such a consensus is the auditing protocol of the IFCN we already mentioned, shared both by journalistic and academic PFC. This stylized verification process has four steps: selection of potentially verifiable claims, selection of evidence sources for verification, scoring process, publication of the score. Let us spell out what happens in each of these steps and how the IFCN justifies its guidelines.

PFC agencies should first identify potentially verifiable claims. Some agencies operate like standard media outlets and set their own agenda of newsworthy events to cover in search of potentially relevant claims. Their staff then analyses relevant speeches and generates a list of potentially verifiable claims. Some agencies also accept claim submissions from their audience. There are two key decisions in this first step: how to distribute the agencies’ attention between the different political parties at work; how to identify the particular claim to be verified. As to the former, the IFCN requires balance: “PFC should not concentrate its fact-checking on any one side” (2.2) As to the latter, the IFCN recommends considering “the reach and importance” of the selected claims and explain the reason for the selection.

Regarding the selection of evidence, the second step in the process, the IFCN recommends the use of publicly available and properly referenced primary sources of evidence. If possible, there should be more than one of those sources and potential conflicts must be disclosed. The IFCN considers the possibility of conflicting evidence about a claim but provides no recommendation to weigh those sources.

As to the third and fourth steps, the scoring process and publication, the IFCN offers even less information, although both processes are methodologically controversial. A couple of examples can serve to illustrate some of the problems. Consider first, the scoring process. The scoring scales of three leading international agencies (Table 1) show no correspondence between them, except for two points: True and False.

Table 1: Truth scales

Scor	Newtral	Politifact (US)	Pagella Politica
1	True	True	True
2	Half true	Almost true	Almost true
3	Misleading	Half true	Not clear
4	False	Mostly false	False
5	-	Fals	Crazy story
6	-	Pants on fire	-

Moreover, the published scoring guidelines of each of these agencies include only broad guidelines that leave ample discretion to the scorer in deciding how to classify a claim, e.g., half-true or mostly false. The problem is further complicated when the claim selection process is considered because abstracting away from a statement’s context or smoothing the verbal nuances may affect the score more or less depending on the scale used.

Controversies over the fourth stage in the verification process, the publication of the verified claims, result from the fact that PFC agencies do not follow a consistent rule about the release of their checks. Rather they often behave like conventional media where publication ultimately depends on editorial decisions. Without a complete record of all the verification activity of an agency, it is impossible to ascertain whether there is any flaw or bias in the fact-checking process.

All in all, the IFCN protocol is just a vague and incomplete set of guidelines that leaves ample room for checkers' discretion. This is surprising since, according to the IFCN, the overarching epistemic principle the protocol should foster is reproducibility. Its audited agencies should "provide all sources in enough detail that readers can replicate their work" and eventually correct the score<sup>6</sup>. There is an entire section of the guidelines on the need to guarantee an "open and honest correction policy", so that an agency's audience can revise the published scores and keep a public record of any past mistakes.

The reproducibility principle would make PFC more an academic than a journalistic enterprise. It is clearly inspired on the ideal of experimental reproducibility: the ability to obtain "the same results from the conduct of an independent study whose procedures are as closely matched to the original experiment as possible."<sup>7</sup> Experimental reproducibility is indeed one of the pillars that grounds the self-correcting nature of science: a failed replication may signal a problem in an already accepted experimental outcome, prompting the community to revise it.

The reproducibility of an experimental outcome is, *prima facie*, an index of its objectivity, in the sense that the intervention under study does not depend on the interests or preferences of the experimenter (Norton, 2015). These interests/preferences are potential sources of *bias*, i.e., systematic deviations from the outcome that would otherwise obtain. If in a clinical trial, the experimenter is allowed to choose to which patients she can assign the treatments under analysis, her interests can contaminate the treatment comparison - e.g., if, out of compassion, she assigns the experimental treatment to the most desperate

---

<sup>6</sup> The relevant passage says: "Signatories want their readers to be able to verify findings themselves. Signatories provide all sources in enough detail that readers can replicate their work, except in cases where a source's personal security could be compromised. In such cases, signatories provide as much detail as possible."

<sup>7</sup> Although the IFCN uses replicability, the definition it uses is consistent with the way reproducibility is understood in the metaresearch literature (Goodman et al., 2018),. It would make no sense, for instance, to speak of the *direct* or *indirect* replication of a fact-checking protocol, given that these protocols are not tracking causal interventions, like most scientific experiments do -see section 4 below.



patients. Hence, treatment randomization serves as a debiasing method. Debiasing methods foster reproducibility, ensuring that experiments do not depend on the experimenter's preferences (Teira, 2016).

Potential biases may affect each stage of the fact checking process. Consider just a few examples. Selecting a claim is not just assessing its relevance and scope. Fact-checkers also need to agree on how to identify a claim, which will have consequences for its truth score. For example, fact-checkers sometimes merge multiple statements into a single claim (or break one into different factual assertions) without laying out the compositional semantics they are using (Uscinski & Butler, 2013).

Consider the claim “Le Pen and Podemos voted in the European parliament against the liberal proposal to recognize Guaido” [as president of Venezuela, our translation]. The Spanish fact-checker Newtral takes this as a single claim assigning a “Misleading” truth score. But, in fact, Newtral's subsequent analysis shows that there are two separate claims: whether the liberals registered the proposal and which parties voted against it, each of them with a different truth value. As to the former, three other parties co-registered the proposal; as to the latter, some other parties also voted against it. But Newtral does not explain under which logic these two sentences are evaluated and combined.<sup>8</sup>

Or take claims that involve implicit statistical data. In the best possible scenario, the verification of such claims would use a reliable statistical source to fact-check every claim. But how should these implicit statistics be interpreted (e.g., relative or absolute changes)? For instance, the Spanish agency Newtral fact-checked the two following claims made by the regional president of Andalusia: “We [Andalusia] have the same number of graduates as countries like Germany” was awarded a “True” score because it was interpreted in percentual terms; “Andalusia is the region that has created more jobs in recent years” was also awarded a true, but it is now interpreted in absolute terms.<sup>9</sup> At all these stages, the personal preferences or interests of the fact-checker may interfere

---

<sup>8</sup> See <https://www.newtral.es/albert-rivera-frente-a-la-propuesta-de-los-liberales-para-reconocer-a-guaido-en-el-parlamento-europeo-votaron-en-contra-le-pen-y-podemos/20190206/> (Last visited on July 21, 2023)

<sup>9</sup> See: <https://www.newtral.es/susana-diaz-tenemos-el-mismo-numero-de-egresados-que-paises-como-alemania/20181130/> and <https://www.newtral.es/susana-diaz-asegura-que-andalucia-es-la-region-que-mas-empleo-ha-creado-en-los-ultimos-anos/20181127/> (Last visited on July 21, 2023)

with the verification process in ways that hinders reproducibility. Thus, if a different fact-checker conducts again the check, the outcome may be different (Lim, 2018).

However, the only debiasing method the IFCN considers in its guidelines is the use of conflict of interest policies. Two of the five sections of its code provide standard recommendations against conflict of interest: disclosure, abstention rules, etc. The IFCN seems to assume that controlling for conflicts of interest will be enough to guarantee reproducibility<sup>10</sup>.

It is open to debate, however, whether conflicts of interest policies can be this effective epistemically (de Melo-Martín & Intemann, 2009). But let us assume, for the sake of the argument, that these policies work appropriately and consider whether PFC achieves some reproducibility and whether it fosters accountability.

Before tackling these questions, we need to introduce a conceptual distinction between *rules* and *standards*. According to some prominent legal scholars (Kaplow, 1992) (Andreoletti & Teira, 2019) rules are prescriptions with a conditional form, in which, if the antecedent obtains, an action should follow. Take, for instance, the IFCN prescription of avoiding a one-sided coverage. A fairness-in-coverage rule to implement would be “If a claim by a politician of ideology X has been verified, consider also a claim by a politician of ideology Y”. In a two-party political system, the implementation of the rule would lead to a perfect 50% balance. The IFCN has instead chosen to articulate its prescription as a *standard*, i.e., a broad guideline hinging on contextual interpretation: PFC should not “concentrate its fact-checking on any one side”. Each fact-checker should decide how many verifications each side should receive. An individual fact-checker may do it wisely, of course, but there is no guarantee that if a different PFC agency conducted the process, they would achieve a comparable balance (Lim, 2018).

Reproducibility is served better by rules than standards. One central debate in the philosophy of replication is the so-called *experimenter’s regress* (Fidler & Wilcox, 2018).

---

<sup>10</sup> A reviewer suggests an alternative interpretation though: conflict of interest rules aim to secure a *fair coverage*, rather than reproducibility. Avoiding a conflict of interest would “screen out biases that could affect the overall picture of political discussions an agency creates over time”, e.g., keeping a balance in verification between the different parties, selecting claims according to their reach and importance, etc. Fair coverage would be, in fact, independent of reproducibility. We agree that this is a plausible interpretation of the conflict of interest rules, but still, readers of the IFCN code are left wondering how the guidelines will foster reproducibility. In any case, this does not affect our claim since conflict of interest rules are also said to contribute to reproducibility, via bias correction, in various experimental disciplines. For clinical trials see, for instance, (Lundh et al., 2017).

To reproduce an experimental result, experimenters draw on a protocol with instructions about how to conduct it. But, after (Star & Collins, 1988), sociologists of science have argued that, since protocols are necessarily incomplete, there is also tacit knowledge about each experimental setup that cannot be made fully explicit and is passed along through personal contact between experimenters. Hence, concludes Collins, there is a subjective dimension to scientific experiments that cannot be eliminated. Whereas rules require minimal interpretation, standards make ample room for tacit knowledge. We have defended elsewhere (Andreoletti & Teira, 2019) that a rule-based approach to scientific experimentation minimizes the consequences of the experimenter's regress, precisely because they foster replicability<sup>11</sup>. However, the preliminary evidence about bias in PFC suggests that PFC agencies in the US do not have very precise verification rules –for instance, they understand balance in very different ways.<sup>12</sup>

In the next section, we are going to discuss how PFC can enhance its reproducibility, to address, in section 5, whether this enhanced reproducibility actually improves the role of PFC in fostering electoral accountability.

#### **4. A coherentist approach to PFC**

Reproducibility is not enough to grant a superior epistemic status to any activity. As (Norton, 2015) correctly argues, in science no amount of replications would persuade many of us to accept that, e.g., patients may improve thanks to other people's prayers. Various forms of (Christian) intercessory prayer have been tested for clinical benefits in a number of trials, sometimes reaching small positive results (Hodge, 2007). According to Norton, there are two reasons to object to the cogency of such an experiment. On the one hand, there is no scientific agreement as to the causal connection between the intervention and the clinical outcomes under study. On the other hand, there is no standardized manner to conduct the intervention: for instance, people can pray daily, weekly, with various degrees of knowledge of the patient, various degrees of devotion, at different stages of the patients' condition. Therefore, there is no agreement on the correct

---

<sup>11</sup> We do not imply that replicability in science is entirely rule-based, with tacit knowledge playing no role. Yet, the lack of explicit enough experimental protocols has been noted as a relevant factor in the replicability crisis in different disciplines (Andreoletti, 2020).

<sup>12</sup> There is an emerging literature about potential biases in PFC, namely about how balanced is the attention they pay to different political parties. The most prominent among these would be a differential treatment of politicians according to their ideology. Although the evidence is still preliminary, according to the fact-checks of some leading US agencies, right-wing politicians would systematically lie more than their left-wing peers -see (Amazeen, 2016; Farnsworth & Lichter, 2019; Marietta et al., 2015). See also (Fernández-Roldán et al., 2023) for a methodological discussion.

way to carry out the trial, and the observed positive effects can be explained by factors that experimenters may simply are not controlling for.

In other words, in science replicability is only epistemically persuasive to the extent that there is a pre-existing agreement in a community of experimenters about (i) the structure of the intervention; (ii) the confounders to be controlled for in the test. In PFC, of course, such agreements are simply out of reach: there is nothing like a causal structure guiding the verification process discussed in the previous section. It is rather a set of guidelines orienting individual decisions based primarily on rather subjective assessments: e.g., the “the reach and importance” of the selected claims.

However, it may be argued that constraining such individual choices in a systematic manner may pave the way for a more scientific fact-checking. A way that those PFC who see themselves as academics are perhaps trying to follow. We may ground such an argument on Hasok Chang’s acclaimed philosophical analysis of the history of thermometry (Chang, 2004). Like fact-checking, thermometry emerged as a practical discipline without any solid theoretical underpinning. It was about using scales to quantify the temperature of different phenomena. However, there was no initial agreement among the practitioners about what temperature was or even about the scale calibration. According to Chang, the disunity of thermometry was overcome thanks to a shared *coherentist* approach. Once an agreement was reached on a concept of temperature with a well-established measurement method for a certain range of phenomena, thermologists accepted any extension of such concept and method to a new domain provided that they were consistent, under certain conditions, with the pre-existing concepts and methods. Gradually, unification was achieved: robust empirical regularities emerged and solid theoretical explanations soon followed.

Perhaps *truth* in political fact-checking is somewhat similar to *temperature*: a multifarious term with apparently discordant interpretations that, nonetheless, may admit a well-established method of measurement in a particular domain -think of simple statistical statements about, e.g., how an economy evolves. If one such method is discovered, it may be gradually extended to other PFC domains leading, at some point, to a proto-scientific discipline. If this expansion occurs within a coherentist framework, reproducibility will play a key role: a consistent extension of a fact-checking method will only be possible if it systematically yields the same outcomes whoever conducts it. Like early thermometry, PFC is today a craft more than a science. But just as the progress of

thermometry into a proto-scientific discipline yielded more efficient industrial processes (e.g., in porcelain manufacturing), some leading practitioners think that more reproducible PFC will foster accountability in our democracies.

How could PFC achieve reproducibility then? The upshot of our previous discussion is that a standard-based approach like the IFCN guidelines is unlikely to yield much agreement in the verification process, since there is no reason to expect that different fact-checkers will make the same decisions at each stage of the process. A well-specified rule-based approach, in principle, should maximize reproducibility, but how far can this approach go? To grasp the challenges involved we should consider the state of the art in automated fact-checking, a discipline nowadays based on Natural Language Processing (NLP) and statistical inference (Vargo et al., 2018; Thorne & Vlachos, 2018; Zeng et al., 2021).

There is indeed a growing interdisciplinary community investigating how to conduct PFC with minimal human intervention, articulating the rules governing each stage of the process and implementing them into software that may conduct it automatically. Following the verification protocols discussed above, let us briefly consider the challenges arising at each stage. First, the identification of candidate claims for verification requires a set of claim categories about which truth can be established using publicly accessible data. One possible division of potentially verifiable claims is between quantitative claims, causal claims, legal claims and predictions (Lazarski et al., 2021) . For each candidate claim, an algorithm would extract, e.g., a semantic representation and check whether it fits in any of these four categories. If it does, another algorithm should extract its content to verify it in a comparison with the relevant *knowledge base*. To extract the content, there are also different options: e.g., to decompose its syntactic structure into a standardized NLP format.

To verify the analysed claim, it is then necessary to identify a relevant source of evidence (the *knowledge base*). Again, there are different options in the literature, ranging from the use of pre-existing facts' databases to algorithms directly retrieving the information from Internet search engines (Zeng et al., 2021). The comparison between the analysed claim and the evidence source usually relies on machine learning techniques with, again, different approaches (Guo et al., 2022). E.g., analysing semantic textual similarity with already verified fact-checks, textual entailment from more complex knowledge bases, veracity prediction, etc. Finally, there is one final, crucial, step in the verification process:

deciding how many veracity labels should be used in the output, with the choices ranging from the simplest *true/false/undetermined* to the fine-grained scales now in use in most PFC agencies (see table 1, above).

Automated fact-checking algorithms would then bring about rule-based reproducibility, at least relative to a knowledge base and a set of algorithms, but with clear limitations (Graves, 2018). Following up on the analogy with temperature, we are still at a stage where there are multiple competing algorithms purportedly measuring the same property but with different, and often discordant, tools: e.g., most algorithms now in use do not process the semantic content of the knowledge base so, on their own, they cannot ascertain the connection of the words with the real world. From the standpoint of coherence, reproducible PFC may be vindicated as a long-term project. As of today, the efficacy of these algorithms is limited mostly to straightforward, declarative statements in English for which a consensual knowledge base exists (e.g., statistical claims). Progress in this field would require some fundamental agreement among researchers on: a) the selection criteria for candidate claims to be verified; b) the choice of the relevant knowledge base; and c) the truth scale on which PFC should focus. Algorithms alone are unlikely to solve any of these questions.

Whatever the future of automated fact-checking, if reproducibility is the epistemic norm guiding the scientific development of PFC, its most accomplished expression will come from a rule-based implementation of verification protocols, of which we have just provided a preliminary illustration<sup>13</sup>. Now that we have established how far the reproducibility of PFC actually reaches, let us discuss whether this proto-scientific project would serve the original mission of PFC, namely enhancing the accountability of politicians.

---

<sup>13</sup> A reviewer objects to whether the appropriate comparison to assess the reproducibility of PFC should not be with the qualitative social sciences (e.g., history). It has been argued that there are different reproducibility benchmark across disciplines (Leonelli, 2018). In those fields where there is a low degree of control on the environment and statistical tools are rarely used, reproducibility requires, at most, that “any skilled experimenter working with same methods and materials would produce similar results”. This is what standard-based PFC could achieve at its best, assuming that PFC agencies could operate without the practical constraints of traditional journalism (limited resources and tight publication deadlines). But even in those ideal circumstance, as we also argue next in section 5, it is dubious that standard-based PFC would foster political accountability better than standard journalism.

## 5. PFC will not foster accountability better than quality journalism

Let us summarise our analysis so far. We started with a split view among fact-checkers about their job description, whether they were journalists or academics. Journalistic PFC would transform verification into a stand-alone practice. Academic PFC thought they should develop stronger verification protocols. Both of them agreed on the reproducibility of these protocols as the guiding epistemic benchmark of their activity. Our analysis in section 3 showed that there is no reason to expect reproducibility from the IFCN standards that identify journalistic PFC. In section 4, focusing on how to strengthen reproducibility, we discussed how academic PFC could evolve into a rule-based enterprise. To close our analysis, we will now wonder whether perfectly reproducible PFC actually serves its ultimate mission, to foster political accountability.

The intuition about the role of journalists in keeping politicians accountable is usually conveyed by Thomas Jefferson's about public officers becoming "wolves" in a world without newspapers, where citizens lose track of how public affairs are managed. This intuition has been fleshed out in different ways, but perhaps the standard rendition, at least among analytically minded social scientists, goes in terms of *electoral accountability*. On the one hand, politicians are accountable if they can be sanctioned by voters when they do not perform well in office. On the other hand, electoral campaigns may be understood as a selection process in which voters can identify those politicians who they can trust to act on their behalf and elect them. Journalism serves both *sanction-* and *trust-based* electoral accountability (Mansbridge, 2014).

Citizens, politicians, and mass media are modelled as self-interested rational agents to study, e.g., under which conditions the number of informed citizens, and the media coverage of politicians have an impact on the electoral outcome<sup>14</sup>. Voters respond to the perceived competence or actions of politicians, as reported by the press, be it through positive or negative coverage (trust-based and sanction-based accountability, respectively). There is indeed evidence that traditional quality journalism seems to have been effective at fostering electoral accountability -see (Strömberg, 2016) for a general

---

<sup>14</sup> Rational choice models of accountability and the media have some underlying dilemmas. Downs (1957) was the first to show that learning about candidates and their policies is beneficial for voters, but since no individual voter will have the power to alter an election outcome with a ballot, it is rational for each one of them not to invest in any learning. Publishing the information required for rendering politicians accountable can be thus conceived as a problem of privately providing a public good, for which solutions exist only under a limited range of circumstances (Bruns & Himmler, 2016).

overview. The question is whether PFC, in any of its forms, may improve upon this benchmark. Our analysis in the previous sections allows us, at least, a preliminary exploration of how rule- or standard- based reproducible PFC may do it.

Let us start with the former: as it stands, a rule-based, quasi-automated, approach to PFC would only be effective in targeting a small subset of statements in political discourse - straightforward sentences in English with a clear knowledge base for the verification. PFC would increase electoral accountability if voters cared for these straightforward lies (or truths), but, if they did, politicians may protect themselves against verification, hedging their sentences with verbal nuances injecting ambiguity. As of today, it is as simple as adding a proviso like “reportedly” before any factual piece of information (Thorne & Vlachos, 2018, p. 8). The epistemic virtue of reproducible fact-checking is also its political weakness: a rule-based approach to verification is not strategy-proof. Politicians will find out how the rules work and arrange their discourse accordingly to avoid fact-checkers<sup>15</sup>. Moreover, politicians might adopt this strategy not necessarily to lie deliberately, but rather to avoid the reputational penalty of a negative score arising from an accidental mistake in a public statement. In this way, PFC could also reduce accountability, as representatives would become too vague in their interventions. The incentives to avoid inaccuracies would result in a worse informed citizenry. Hence, our first conclusion: if academic PFC should be a rule-based enterprise for the sake of reproducibility, it is dubious that it will foster electoral accountability, as compared to traditional quality journalism.

This vulnerability of the rule-based approach may explain why the IFCN opted for a set of guidelines articulated on interpretable standards. Historically, politicians do not seem to indulge in straightforward lies, at least in liberal democracies. Traditional investigative journalists require significant budgets to gain access to the relevant documentation and witnesses. There are no general rules for handling this information and there is no algorithmic recipe to report it. For each story, the investigative journalist should use her

---

<sup>15</sup> This is, of course, nothing but a conjecture, theoretically inspired by the law and economics analysis of *avoidance*. Criminals engage in avoidance activities to reduce the probability of punishment or its magnitude: covering up incriminating evidence or abusing evidentiary rules and procedures. Against a naive view of accountability, law & economics scholars have shown that, under standard economic rationality, increasing punishment, instead of deterring criminals, may lead them to more avoidance (Nussim & Tabbach, 2009) In our case, the cost for a politician to avoid a rule-based fact-checker is so low (a mere play on words) that we may safely assume they will give it a try –rather than facing the electoral penalties of being caught lying.



own judgment in deciding how to persuade the reader that the evidence provided proves a claim true or false.

Investigative journalism is sometimes called *accountability reporting* (Knobel, 2018): unlike in automated PFC, it is not easy for politicians to get away with their lies if a well-funded research team is tasked to check them. At the same time, the audience needs to trust the journalists to do their job in an unbiased manner, since there are no general rules to check whether the reporting is correct. Readers must validate the journalists' methodological judgments on their own.

Our second question is: can standard-based, journalistic PFC increase accountability more than traditional quality journalism? In a world in which there are fewer resources for investigative journalism and less trust in the journalists' impartiality, the IFCN guidelines are clearly trying to strike a compromise. Their interpretable standards leave room for individual judgment in verification, but they should be specific enough as to allow readers to reproduce the fact-check, enhancing their trust in the process. Following these guidelines, PFC would intuitively foster electoral accountability: citizens' trust in reproducible PFC would lead them to orient their votes according to their verifications. The answer to our second question is also negative: there is no evidence so far that journalistic PFC fosters accountability<sup>16</sup>, quite expectable given the problems we detected in the IFCN protocol, in section 3. It seems as if audiences are not trusting PFC much. Our conclusion is that, for PFC, there is no way out. Rule-based reproducibility will not foster accountability. Standard-based, journalistic, PFC loses reproducibility without delivering any clear accountability gains, as compared to traditional journalism.

Still, it may be argued that PFC does no harm either. It is just a complement to other forms of journalism and it may be sometimes helpful. Time will tell, but we want to close our analysis with a warning about a potentially harmful side-effect of PFC. How it confuses audiences presenting their truth-scales as if they were scientific measurement instruments (as in Politifact's *Truth-o-meter*<sup>17</sup>), when in fact they are uncalibrated *survey tools* (Johnson & Morgan, 2016).

The illusion behind many truth-scales is that they appear to be like thermometers: using them according to the instructions should yield, time and again, the same measurement

---

<sup>16</sup> Experimental research suggests that measurable effects on vote change are small (Nyhan et al., 2020)

<sup>17</sup> See <https://www.politifact.com/truth-o-meter/>

outcome, with an established margin of error. In the analogy, the truth of the proposition would be, so to speak, the causal force yielding the truth score, just as thermometer readings are ultimately caused by movement of the molecules in its fluid. But the truth scales sponsored by the IFCN signatories are not thermometers, but rather simple *questionnaires*. For each claim a politician utters, the audience asks the fact-checker which score grasps better its truth, under the assumption that the fact-checker will follow a protocol to provide the correct answer. But our fieldwork reveals that these protocols leave ample room for the fact-checker to choose almost any answer. What the score reveals is just *her personal, qualified view about the claim*.

Let us illustrate this concern with our fieldwork at Newtral, a Spanish PFC, that uses a four-point truth scale approved by the IFCN -for a full-fledged discussion see (Fernández-Roldán et al., 2023). For Newtral, True and False are self-explanatory labels, but its two intermediate scores are explained as follows: “Half-true” and “Misleading” apply to “correct” statements that may be erroneously interpreted as true or false depending on additional information about: (a) its context, (b) the data it contains, and (c) its semantics/pragmatics<sup>18</sup>. We asked Newtral fact-checkers how they made these choices without guidelines and the answer was that they followed their own judgment. The score expresses this judgment, rather than the claim’s truth value. Interestingly, a third of the fact-checks released by Newtral between October 2018 and 2019 had one of these intermediate scores.

Why should we assume that any of these four scores grasps the truthfulness of a claim? What sort of information about a claim should drive the scoring process? Newtral does not provide an answer to these questions. And reproducibility means very little in this context: if an audience was presented with the claim and the truth scale alone, there is no guarantee that they would reach the same score or that this score would capture the actual truth of the proposition. We would be just probing how knowledgeable this audience is

---

<sup>18</sup> Newtral’s truth scores and methodology are accessible here: <https://www.newtral.es/metodologia-transparencia/> (Last visited on July 21, 2023). The truth scores are described as follows: “True: the claim is rigorous and there is neither context nor relevant additional data missing. Half true: the claim is correct, although it needs clarification, additional information or context. Misleading: the claim contains correct data, but neglects very relevant elements (*sic*) or mixes incorrect data conveying an impression different (*sic*), imprecise or false. False: the claim is false” (Our translation, the highlighted bits are not grammatical in the original Spanish either). Newtral has been positively audited by the IFCN 6 times over the last 7 years: <https://ifcncodeofprinciples.poynter.org/profile/newtral> (Last visited on July 21, 2023).

about such claim in comparison with the Newtral fact-checking team: for instance, whether the audience is able to interpret the claim in the same way as Newtral, whether the audience has the relevant background information that Newtral has used, etc. And, given how loose PFC agencies are about their scoring methodology (e.g., how to tell apart “half-truths” from “misleading statements”), even if the audience reached the same score than the PFC agency, it may well be a random coincidence without epistemic merit.

To conclude, there is no reason to expect that PFC, journalistic or academic, will foster electoral accountability better than traditional quality journalism. But they may undermine the audience’s trust in journalistic objectivity by presenting their truth scale as thermometers, when they are, in fact, multiple choice questions about political claims, without a clear procedure to answer them. PFC simply asked the audience to believe the score, justified, at best, by a short paragraph summarizing the supporting evidence. This is certainly cheap and easy to digest, but we wonder whether liberal democracies would not be better off with traditional quality journalism or contextual corrections as other fact-checkers do (Uscinski, 2015): present as thoroughly as possible the information a journalist can obtain about each case, and let the audience make up their minds.

## References

- Amazeen, M. A. (2016). Checking the Fact-Checkers in 2008: Predicting Political Ad Scrutiny and Assessing Consistency. *Journal of Political Marketing*, 15(4), 433–464. <https://doi.org/10.1080/15377857.2014.959691>
- Amazeen, M. A., Thorson, E., Muddiman, A., & Graves, L. (2018). Correcting Political and Consumer Misperceptions: The Effectiveness and Effects of Rating Scale Versus Contextual Correction Formats. *Journalism and Mass Communication Quarterly*, 95(1), 28–48. <https://doi.org/10.1177/1077699016678186>
- Andreoletti, M. (2020). Replicability Crisis and Scientific Reforms: Overlooked Issues and Unmet Challenges. *International Studies in the Philosophy of Science*, 33(3), 135–151. <https://doi.org/10.1080/02698595.2021.1943292>
- Andreoletti, M., & Teira, D. (2019). Rules versus Standards: What Are the Costs of Epistemic Norms in Drug Regulation? *Science Technology and Human Values*, 44(6), 1093–1115. <https://doi.org/10.1177/0162243919828070>
- Birks, J. (2019). *Fact-Checking Journalism and Political Argumentation: A British Perspective*. Palgrave. <https://doi.org/10.1007/978-3-030-30573-4>

- Braun, J. A., & Eklund, J. L. (2019). Fake News, Real Money: Ad Tech Platforms, Profit-Driven Hoaxes, and the Business of Journalism. *Digital Journalism*, 7(1), 1–21. <https://doi.org/10.1080/21670811.2018.1556314>
- Breen, M., & Gillanders, R. (2020). Press freedom and corruption perceptions: Is there a reputational premium? *Politics and Governance*, 8(2), 103–115. <https://doi.org/10.17645/pag.v8i2.2697>
- Bruns, C., & Himmler, O. (2016). Mass media, instrumental information, and electoral accountability. *Journal of Public Economics*, 134, 75–84. <https://doi.org/10.1016/j.jpubeco.2016.01.005>
- Carson, A., Gibbons, A., Martin, A., & Phillips, J. B. (2022). Does Third-Party Fact-Checking Increase Trust in News Stories? An Australian Case Study Using the “Sports Rorts” Affair. *Digital Journalism*, 10(5), 801–822. <https://doi.org/10.1080/21670811.2022.2031240>
- Chang, H. (2004). *Inventing Temperature: Measurement and Scientific Progress*. Oxford University Press. <https://doi.org/10.1093/0195171276.001.0001>
- de Melo-Martín, I. & Intemann, K. (2018). *The Fight Against Doubt: How to Bridge the Gap Between Scientists and the Public*. Oxford University Press. <https://doi.org/10.1093/oso/9780190869229.001.0001>
- de Melo-Martín, I. & Intemann, K. (2009). How do disclosure policies fail? Let us count the ways. *The FASEB Journal*, 23(6), 1638–1642. <https://doi.org/10.1096/fj.08-125963>
- Downs, A. (1957). An Economic Theory of Political Action in a Democracy. *Journal of Political Economy*, 65(2), 135–150. <https://doi.org/10.1086/257897>
- Elías, C. (2019). *Science on the Ropes*. Springer. <https://doi.org/10.1007/978-3-030-12978-1>
- Elliott, K. C. (2017). *A Tapestry of Values: An Introduction to Values in Science*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780190260804.001.0001>
- Farnsworth, S., & Lichter, R. (2019). Partisan Targets of Media Fact-checking: Examining President Partisan Targets of Media Fact-checking: Examining President Obama and the 113th Congress. *Virginia Social Science Journal*, 53, 51–62.
- Fernández-Roldán, A., Elías, C., Santiago-Caballero, C., & Teira, D. (2023). Can We Detect Bias in Political Fact-Checking? Evidence from a Spanish Case Study. *Journalism Practice*. <https://doi.org/10.1080/17512786.2023.2262444>
- Fidler, F., & Wilcox, J. (2018). Reproducibility of Scientific Results. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab - Stanford University.

- Galison, P. (2015). The Journalist, the Scientist, and Objectivity. In Flavia Padovani, Alan Richardson, & Jonathan Y. Tsou (Eds.), *Objectivity in Science* (pp 57-75). Springer. [https://doi.org/10.1007/978-3-319-14349-1\\_4](https://doi.org/10.1007/978-3-319-14349-1_4)
- Gans, H. J. (2004). *Democracy and the News*. Oxford University Press.
- Graves, L. (2016). *Deciding what's true: The rise of political fact-checking in American journalism*. Columbia University Press.
- Graves, L. (2018). Boundaries Not Drawn: Mapping the institutional roots of the global fact-checking movement. *Journalism Studies*, 19(5), 613–631. <https://doi.org/10.1080/1461670X.2016.1196602>
- Graves, L., & Cherubini, F. (2016). *The Rise of Fact-Checking Sites in Europe*. Reuters Institute -University of Oxford. <https://reutersinstitute.politics.ox.ac.uk/our-research/rise-fact-checking-sites-europe>
- Graves, L., & Konieczna, M. (2015). Sharing the news: Journalistic collaboration as field repair. *International Journal of Communication*, 9(1), 1966–1984.
- Guo, Z., Schlichtkrull, M., & Vlachos, A. (2022). A Survey on Automated Fact-Checking. *Transactions of the Association for Computational Linguistics*, 10, 1–28. [https://doi.org/10.1162/tacl\\_a\\_00454](https://doi.org/10.1162/tacl_a_00454)
- Hansson, S. O. (2018). How connected are the major forms of irrationality? An analysis of pseudoscience, science denial, fact resistance and alternative facts. *Mètode*, 8, 125–131. <https://doi.org/10.7203/metode.8.10005>
- Hodge, D. R. (2007). A systematic review of the empirical literature on intercessory prayer. *Research on Social Work Practice*, 17(2), 174–187. <https://doi.org/10.1177/1049731506296170>
- Johnson, R. L., & Morgan, G. B. (2016). *Survey Scales: A Guide to Development, Analysis and Reporting*. Guilford Press.
- Kaplow, L. (1992). Rules versus standards: An economic analysis. *Duke Law Journal*, 42(3), 557–629.
- Knobel, B. (2018). *The Watchdog Still Barks*. Fordham University Press. <https://doi.org/https://doi.org/10.2307/j.ctt201mpdz>
- Lazarski, E., Al-Khassaweneh, M., & Howard, C. (2021). Using NLP for fact checking: A survey. *Designs*, 5(3), 1-22. <https://doi.org/10.3390/designs5030042>
- Leonelli, S. (2018). Rethinking reproducibility as a criterion for research quality. *Research in the History of Economic Thought and Methodology*, 36 (B), 129-146 <https://doi.org/10.1108/S0743-41542018000036B009>
- Lim, C. (2018). Checking how fact-checkers check. *Research and Politics*, 5(3), 1–7. <https://doi.org/10.1177/2053168018786848>
- Lundh, A., Lexchin, J., Mintzes, B., Schroll, J. B., & Bero, L. (2017). Industry sponsorship and research outcome. *Cochrane Database of Systematic Reviews*, 2(2), MR000033 <https://doi.org/10.1002/14651858.MR000033.PUB3>

- Mansbridge, J. (2014). A Contingency Theory of Accountability. In M. Bovens, R. Goodin, & T. Schillemans (Eds.), *The Oxford Handbook of Public Accountability* (pp.55-68). Oxford University Press.
- Mantzarlis, A. (2018). Fact-Checking 101. In *Journalism, fake news & disinformation: Handbook for journalism education and training*. UNESCO.  
<https://en.unesco.org/node/296054>
- Marietta, M., Barker, D. C., & Bowser, T. (2015). Fact-checking polarized politics: Does the fact-check industry provide consistent guidance on disputed realities? *The Forum*, 13(4), 577–596. <https://doi.org/https://doi.org/10.1515/for-2015-0040>
- Meyers, C. (2020). Partisan News, the Myth of Objectivity, and the Standards of Responsible Journalism. *Journal of Media Ethics: Exploring Questions of Media Morality*, 35(3), 180–194. <https://doi.org/10.1080/23736992.2020.1780131>
- Norton, J. D. (2015). Replicability of experiment. *Theoria*, 30(2), 229–248.  
<https://doi.org/10.1387/theoria.12691>
- Nussim, J., & Tabbach, A. D. (2009). Deterrence and avoidance. *International Review of Law and Economics*, 29(4), 314–323. <https://doi.org/10.1016/j.irl.2009.05.001>
- Nyhan, B., Porter, E., Reifler, J., & Wood, T. J. (2020). Taking Fact-Checks Literally But Not Seriously? The Effects of Journalistic Fact-Checking on Factual Beliefs and Candidate Favorability. *Political Behavior*, 42(3), 939–960.  
<https://doi.org/10.1007/s11109-019-09528-x>
- Stalph, F. (2018). Truth Corrupted: The Role of Fact-Based Journalism in a Post-Truth Society. In H. Oliver & S. Florian (Eds.), *Digital Investigative Journalism: Data, Visual Analytics and Innovative Methodologies in International Reporting* (pp. 237-248). Palgrave Macmillan. [https://doi.org/10.1007/978-3-319-97283-1\\_22](https://doi.org/10.1007/978-3-319-97283-1_22)
- Star, S. L., & Collins, H. M. (1988). *Changing Order: Replication and Induction in Scientific Practice*. The University of Chicago Press.  
<https://doi.org/10.2307/3105095>
- Stencel, M., & Griffin, R. (2018, February 22). Fact-checking triples over four years. In *Duke Reporters' Lab*. <https://reporterslab.org/fact-checking-triples-over-four-years/>
- Stencel, M., & Luther, J. (2020, October 13). Fact-checking count tops 300 for the first time. In *Duke Reporters' Lab*. <https://reporterslab.org/fact-checking-count-tops-300-for-the-first-time/>
- Stencel, M., Ryan, E., & Luther, J. (2022, June 17). Fact-checkers extend their global reach with 391 outlets, but growth has slowed. In *Duke Reporters' Lab*.  
<https://reporterslab.org/fact-checkers-extend-their-global-reach-with-391-outlets-but-growth-has-slowed/>
- Strömberg, D. (2016). Media Coverage and Political Accountability: Theory and Evidence. In Simon P. Anderson., Joel Waldfogel., & David Strömberg. (Eds.), *Handbook of Media Economics* (pp. 595-622). North-Holland.

- Teira, D. (2016). Debiasing methods and the acceptability of experimental outcomes. *Perspectives on Science*, 24(6), 722–743. [https://doi.org/10.1162/POSC\\_a\\_00230](https://doi.org/10.1162/POSC_a_00230)
- Thorne, J., & Vlachos, A. (2018). Automated fact checking: Task formulations, methods and future directions. *COLING 2018 - 27th International Conference on Computational Linguistics, Proceedings*. <https://doi.org/https://doi.org/10.48550/arXiv.1806.07687>
- Uscinski, J. E. (2015). The Epistemology of Fact Checking (Is Still Naïve): Rejoinder to Amazeen. *Critical Review*, 27(2), 243–252. <https://doi.org/10.1080/08913811.2015.1055892>
- Uscinski, J. E., & Butler, R. W. (2013). The Epistemology of Fact Checking. *Critical Review*, 25(2), 162–180. <https://doi.org/10.1080/08913811.2013.843872>
- Vargo, C. J., Guo, L., & Amazeen, M. A. (2018). The agenda-setting power of fake news: A big data analysis of the online media landscape from 2014 to 2016. *New Media and Society*, 20(5), 2028–2049. <https://doi.org/10.1177/1461444817712086>
- Vicario, M. Del, Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., Stanley, H. E., & Quattrociocchi, W. (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences of the United States of America*, 113(3), 554–559. <https://doi.org/10.1073/pnas.1517441113>
- Walter, N., Cohen, J., Holbert, R. L., & Morag, Y. (2020). Fact-Checking: A Meta-Analysis of What Works and for Whom. *Political Communication*, 37(3), 350–375. <https://doi.org/10.1080/10584609.2019.1668894>
- Zeng, X., Abumansour, A. S., & Zubiaga, A. (2021). Automated fact-checking: A survey. *Language and Linguistics Compass*, 15(10), 178–206. <https://doi.org/10.1111/lnc3.12438>