Web Literacy for College Students
CONTENTS

Part I. Four Fact-Checking Strategies

1. Why This Book? 3
2. Four Fact-Checking Strategies 4
3. When Emotional Content Goes Viral 5

Part II. Previous Verification

4. Has it Already Been Fact-Checked? 9
5. Fact-checking Sites 12
6. Wikipedia 15

Part III. Locate the Source

7. Find the Source 19
8. Identifying Sponsored Content 23
9. Activity: Spot Sponsored Content 26
10. Understanding Syndication 27
11. Tracking the Source of Viral Content 29
12. Tracking the Source of Viral Photos 34
13. Using Google Reverse Image Search 42
14. Filtering by Time and Place to Find the Original 51
15. Activity: Trace Viral Photos 58
Part IV. Credibility of the Source

16. What “Reading Laterally” Means
17. Evaluating a Website or Publication’s Authority
18. Basic Techniques: Domain Searches, WHOIS
19. Activity: Evaluate a Site
20. Time-Saving Tips for Fact-Checkers
21. Finding a Journal’s Impact Factor
22. Using Google Scholar to Check Author Expertise
23. How to Think about Research
24. Finding High Quality Secondary Sources
25. Choosing Your Experts First
26. Evaluating News Sources
27. What Makes a Trustworthy News Source?
28. National Newspapers of Record
29. Activity: Expert or Crank?
30. Activity: Find Top Authorities for a Subject

Part V. Field Guide

31. Verifying Twitter Identity
32. Activity: Verify a Twitter Account
33. Using the Wayback Machine to Check for Page Changes
34. Finding Out When a Page Was Published Using Google
35. Citation Rates
36. Using Google Books to Track Down Quotes
37. Understanding Astroturf
38. Searching TV Transcripts with the Internet Archive
39. Treating Google’s “Snippets” with Suspicion
40. Using Buzzsumo to Find Highly Viral Stories
PART I
FOUR FACT-CHECKING STRATEGIES
WHY THIS BOOK?

When Tim Berners-Lee invented the web over thirty years ago, he envisioned an “open platform that would allow everyone, everywhere to share information, access opportunities and collaborate across geographic and cultural boundaries”¹. In many ways, his vision has been realized; however, in the last few years he is concerned by three trends that are preventing the web from reaching its true potential.

One of the trends that Berners-Lee (World Wide Web Foundation, 2017) notes is the ease with which misinformation can spread on the web. Today, many of us get our information from social media sites and search engines. The very nature of these sites allows for the rampant spread of fake news and misinformation. Students need concrete strategies for tracing claims to sources and for analyzing the reliability of those sources.

The web gives us many strategies, tactics, and tools which can get students closer to the truth of a statement or image within seconds. The web is both the largest propaganda machine ever created and the most amazing fact-checking tool ever invented. But if we have not taught our students those fact-checking capabilities, is it any surprise that propaganda is winning?

This guide will show you how to use date filters to find the source of viral content, how to assess the reputation of a scientific journal, and how to see if a tweet is really from a famous person or an impostor. It will show you how to find pages that have been deleted and figure out who paid for the website you are reviewing.

You will learn how to check a *Wikipedia* page for recent vandalism and how to search the text of almost any print book to verify a quote. This guide will teach you to parse URLs and scan search result blurbs so that you are more likely to get to the right result on the first click. It will show you how to avoid baking confirmation bias into your search terms.

In other words, this guide will help you become “web literate” by showing you the unique opportunities and pitfalls of searching for truth on the web.

---

¹ (World Wide Web Foundation, 2017, para. 1)
FOUR FACT-CHECKING STRATEGIES

When students are looking for information on the web, there are strategies they can use to help them determine whether the information is accurate.

- **Previous verification** – Check to see if someone has already fact-checked the accuracy of information.
- **Locate the source** – Find the original source of the information or claim to determine its trustworthiness.
- **Credibility of source** – Verify the credentials of the author or creator of the source and read what others have said or written about the source.
- **Start over** – If you are not able to find sources to substantiate the information or claim, look for another source.

If any of these strategies prove to be successful in verifying the content, your work is done.

When you encounter a claim that you want to verify, your first step is to see if any fact-checking sites like Politifact, or Snopes, or even Wikipedia have already verified or disputed the claim. If this proves to be a dead-end, your next step is to locate the source of the information.

If the claim is about research, try to locate the journal it originally appeared in. If the claim is about an event, look for the news publication in which the event was originally reported. If the source of information is reputable, such as the journal Science or the newspaper the New York Times, you can stop here. If not, your next strategy is to find out what others have written about the source.

Look for information about the author of the claim, as well as any credentials and affiliations that they may have. Are there any reviews or commentaries about the information?

If you find that the source of the information is not trustworthy, or cannot be verified at all, then your last strategy is to circle back, and start the process over again. Try finding an alternate source that can be verified.
In addition to the four strategies discussed in the last section, it is especially important to fact-check claims when they evoke strong emotions. Some of the information found on social media or the web may cause us to experience emotions such as happiness, anger, pride or vindication, and our inclination is to immediately share this information with others. Remember to verify these claims, just as you would when you fact-check statistics for inclusion in an academic paper.

Below are some examples of emotionally-charged tweets:

You do not need to know much of the background of this tweet to see its emotionally-charged nature. President Trump had insulted Chuck Schumer, a Democratic Senator from New York, and characterized the tears that Schumer shed during a statement about refugees as “fake tears.” This tweet reminds us that Senator
Schumer’s great-grandmother died at the hands of the Nazis, which could explain Schumer’s emotional connection to the issue of refugees.

Or does it? Do we actually know that Schumer’s great-grandmother died at the hands of the Nazis? And if we are not sure this is true, should we really be retweeting it?

Our normal inclination is to ignore the need to fact-check when we react strongly to content, and researchers have found that content that causes strong emotions (both positive and negative) spreads the fastest through our social networks. For further reading, see “What Emotion Goes Viral the Fastest?” by Matthew Shaer. Activists and advocates take advantage of our emotional response to information.
PART II
PREVIOUS VERIFICATION
HAS IT ALREADY BEEN FACT-CHECKED?

When fact-checking a particular claim, quote, or article, check to see if someone has already done the verification work for you. A reputable fact-checking site or subject wiki may have already successfully traced claims to their source, identified the owners of various sites, and linked to reputable sources for counterclaims.

If the fact checking site is reputable and trustworthy, then you can be confident in using the source or not.

Constructing a Query to Find Previous Fact-Checking

Check for previous fact-checking by searching known and trusted fact-checking sites for a phrase or keyword from the story. Use the “site” function in search engines such as Google and DuckDuckGo.

Site Search. The Site Search function allows you to search a specific domain(s) for a search term instead of the entire internet.

In this example, we used the DuckDuckGo search engine. For our query, we use Politifact and Snopes (well-known fact-checking sites) as well as keywords from the story in the following syntax:

obama iraqi visa ban 2011 site:snopes.com site:politifact.com
Here are the results of our search:

A set of DuckDuckGo search results. The top results are from fact-checking sites Snopes and Politifact.

The results show that work has already been done in this area. In fact, the first result from *Snopes* answers our question almost fully. Remember to follow search engine best practice – scan the results and click on the best result in the result list. There are similar syntaxes you can use in *Google*, but for various reasons this particular search is easier in *DuckDuckGo*.

Let’s look at another claim, this time from the President. This claim is that police officer deaths increased 56 percent from 2015 to 2016. Here it is in context:

> can to help you meet those demands. That includes a zero tolerance policy for acts of violence against law enforcement. We all see what happens. We all see what happens and what’s been happening to you. It’s not fair.

> We must protect those who protect us. The number of officers shot and killed in the line of duty last year increased by 56 percent from the year before. Last year, in Dallas, police officers were targeted for execution -- think of this. Who ever heard of this? They were targeted for execution.

An excerpt of a President Trump speech
Let’s ramp it up with a query that checks four different fact-checking sites:


This gives us back a helpful array of results. The first, from the *Washington Post*, actually answers our question, but some of the others provide some helpful context as well.

DuckDuckGo search results. The top search result is an article from the Washington Post fact-checker and we can see the highlighted text that matches our query.

The *Washington Post* confirms that this claim is true. We do not need to search any further.
5. FACT-CHECKING SITES

Advise from Professional Fact-Checkers

1. Leave the site to research it.
2. Ignore the order of search results in Google.
3. Are the sources cited? Is there documentation?
4. Watch for “click-bait.”
5. Watch for inflammatory language, as well as more subtle forms of persuasion.
6. If the site allows readers to comment, read them.
7. Read multiple sources of information to get a variety of viewpoints and media frames.

Some Reputable Fact-Checking Organizations

The following organizations are generally regarded as reputable fact-checking organizations focused on U.S. national news:

- Politifact
- Factcheck.org
- Washington Post Fact Checker
- Snopes
- Truth be Told
- NPR Fact-Check
- Lie Detector (Univision, Spanish language)
- Hoax Slayer

1. Milne Library, 2017
Respected specialty sites cover niche areas such as climate or celebrities. Here are a few examples:

- Climate Feedback
- SciCheck
- Quote Investigator

There are many fact-checking sites outside the U.S.:

- FactsCan (Canada)
- El Polignafo (Mexico)
- The Hound (Mexico)
- Guardian Reality Check (UK)
- BBC Reality Check (UK)
- Channel 4 Fact Check (UK)
- Full Fact (UK)

At the TEDSalon in London, Markham Nolan shares the investigative techniques he and his team use to verify information in real-time, to let you know if that Statue of Liberty image has been doctored or if that video leaked from Syria is legitimate.²

---

A TED element has been excluded from this version of the text. You can view it online here: https://pressbooks.nscc.ca/webliteracy/?p=35
When researching a topic, *Wikipedia* should be approached with caution, especially with articles that are covering contentious subjects or evolving events; however, it is a useful source to get a consensus viewpoint on a subject. Because the *Wikipedia* community has strict rules about sourcing facts to reliable sources, and because authors must adopt a neutral point of view, its articles are often the best available introduction to a subject on the web.

If you can find a claim expressed in a *Wikipedia* article, you can almost always follow the footnote for the claim to a reliable source. Scholars, reporters, and students can all benefit from using *Wikipedia* to quickly find authoritative sources for claims.

For example, the **Canada–U.S. softwood lumber dispute** is one of the largest and most enduring trade disputes between both nations. The heart of the dispute is the claim that the Canadian lumber industry is unfairly subsidized by federal and provincial governments, as most timber in Canada is owned by provincial governments. Canada has won several NAFTA (North American Free Trade Agreement) challenges involving softwood issues, but despite this, the US continues to impose tariffs on Canadian softwood lumber.

At the end of the Wikipedia article, there are links to several authoritative sources. These sources include newspaper articles, Canadian and US government documents, and fact sheets. Using Wikipedia in this fashion is a quick way get to reputable sources on a topic.
PART III
LOCATE THE SOURCE
If we are not able to find evidence that a claim was previously fact-checked, or if the work that was done was insufficient for your purposes, we employ our second strategy and search for the source of the information.

Consider this claim on the conservative site the *Blaze*:

**Report: US Government Ethics director approved controversial Trump tweets**

Controversial tweets from the U.S. Office of Government Ethics that praised President-elect Donald Trump were approved by Director Walter M. Shaub personally, the Daily

**Is this claim true?**

We can check the credibility of this article by considering the author, the site, and the date it was last revised; however, it is of little value to start from the Blaze website. Why? Because like many news pages on the web, this one provides no original information. It is just a rewrite of another news story (see below):
Controversial tweets from the U.S. Office of Government Ethics that praised President-elect Donald Trump were approved by Director Walter M. Shaub personally, the Daily Dot reported Friday.

Through a Freedom of Information Act, the Daily Dot reported that Shaub sent an email ordering an OGE official to post the tweets. The series of tweets posted Nov. 30 applauded Trump for his supposed efforts to alleviate conflicts of interest with his businesses.

The *Daily Dot* is the originator of this news story, and it is this news site that has collected all the information, presumably fact-checked it, and written the article. It is what we call “reporting on reporting” so there is no point in evaluating the *Blaze* article.

So what do we do? We go to the original story and evaluate it. When you go to the *Daily Dot*, you can start asking questions about the credibility of the site or the source. Another good practice is to check out the primary sources used in the Daily Dot news story.

**What are Primary, Secondary and Tertiary Sources**

There are three types of sources of information: primary, secondary, and tertiary. It is important to understand these types and to know what type is appropriate for your coursework prior to searching for information.

**Primary sources**

Primary sources are *original documents that were created during the studied period of time*. They vary a lot from one field of study to another:

- **Humanities**: birth or death registries, censuses, correspondences, historic treaties, interviews, manuscripts, maps, newspapers, notarial acts, novels, paintings, period artifacts, photographs, poems, sculptures, statistics, surveys, testimonies, videos, etc.
- **Sciences**: articles or theses detailing an original study, case notes or observations, clinical exams, conferences, experimental protocols, industrial drawings, lab notes, patents, the periodic table, raw data sets or results, technical reports or forms, etc.

**Secondary sources**

Secondary sources (also called *academic sources or scientific sources*) are analytical documents that interpret
primary sources. Among other things, they include books, electronic resources, memoirs, monographs, peer-reviewed articles, and theses.

Peer-reviewed publications are more reliable than Internet sources (especially when they do not have authors). In fact, Internet sources should be avoided, except for government websites, websites of reputable associations or websites approved by the instructor.

**Tertiary sources**

Tertiary sources are descriptive documents that compile or index primary and secondary sources. Among other things, they include bibliographies, encyclopaedic articles, indexes and registers, library catalogues, and specialized databases. Generally speaking, tertiary sources aid in the identification of sources rather than the evaluation of their content.¹

---

1. (University of Ottawa, 2016)
IDENTIFYING SPONSORED CONTENT

With sponsored content, it is particularly important to locate the source of the claim before evaluating it. For example, review the headlines (below) from a popular technology magazine:

Look at the headline in the upper left corner. Are lawmakers really concerned about this insane military scope? Maybe. But *Network World* is not making this claim (below). Note that the ZeroTac Tactical Scope company is making the claim. It is an advertisement served from another site into this page in a way that makes it look like a story.
Sometimes sponsored content contains helpful information. For example, the piece below, provides an in-depth look at some current industry trends in information technology.
The source of this article is not *InfoWorld*, but the technology company Hewlett Packard, and the piece is written by a Vice President of Hewlett Packard, with no *InfoWorld* oversight. Sponsored content can be a challenge to spot, so make sure you look at the fine print above or below the article.

After you leave college and begin your career, you may be asked by your employer to find information on a topic in the fields of business, technology, or health. For example, your supervisor needs to determine whether a particular piece of very expensive office equipment would aide in streamlining a workplace process. Your job is to find articles that show whether introducing this piece of equipment will be a benefit to the office. You want unbiased information that gives both the pros and cons of introducing the equipment into the workplace. You do not want vendor-biased content at this stage of your information search.

If your employer decides to purchase the equipment, then you will investigate the various vendors and compare costs, features, service agreements, reputations, etc.

**Additional Reading**

3 Ways to Identify Sponsored Content

  Disclosing Branded Content: What You Need to Know
ACTIVITY: SPOT SPONSORED CONTENT

Rank the following news sources on how much sponsored content you believe their pages will feature: CNN, Buzzfeed, Washington Post, HuffPost, Brietbart, New York Times.

Individually, or in groups, visit the following pages and list all sponsored content you see, tallying up the total amount on each page. Then rank the sites from most sponsored content to least.

Rubric Needed for Activity


After you’ve ranked the websites, answer these questions:

1. Did the ranking surprise you at all?
2. What do you think the quantity of sponsored content indicates about a website?
3. How does this change your perspective on these websites’ reliability?
4. Why would some websites have more sponsored content than others?
10.

UNDERSTANDING SYNDICATION

Dictionary.com defines a “syndicate” in journalism as an agency that buys articles, stories, columns, photographs, comic strips, or other features. It distributes them for simultaneous publication in a number of newspapers or periodicals in different localities.

Consider this New York Times web page:

![Screenshot of New York Times](image)

In the left column (above), there are a set of stories – “Germany’s Latest Best Seller?” & “Isis Claims Responsibility” written by New York Times staff.

In the middle column, there are stories – “UK Stock Market Hits Record” & “The Latest: Lebanon Bids Farewell to Istanbul Victims” that are identified as being from the Associated Press and Reuters. You click through to a page that’s on the New York Times site, but not by the New York Times.
To evaluate the source of these articles, your evaluation will not focus on the *New York Times*. You will focus on the reporting record of the *Associated Press* and *Reuters*.

For example, if you are a reader of online news, you may notice that an article critical of a politician or policy suddenly disappears from a news site. The story that disappeared is most likely a syndicated article. *Associated Press* articles, for example, are displayed on the site for a few weeks, then “roll off” and disappear from the site. Why? Because the *New York Times* only pays the *Associated Press* to show them on the site for a few weeks.

You will also occasionally see people complaining about a story from the *New York Times*, claiming it shows a New York “liberal bias” only to find the story was not even written by the *New York Times*, but by the *Associated Press*, *Reuters*, or some other syndicator.

Finding the source means following a piece of content to its true source, and beginning your analysis there. Your first question when looking at a claim on a page should be “Where did this come from, and who produced it?”
TRACKING THE SOURCE OF VIRAL CONTENT

In our previous examples, it has not been too difficult to trace the source of the information or claim. The Blaze story, clearly links to the Daily Dot piece so that anyone reading their summary is one click away from confirming it with the source. The New York Times makes it apparent that the syndicated content is from the Associated Press and Reuters, so checking the credibility of the sources is fairly simple.

This is good internet citizenship. Articles on the web that re-purpose other information or artifacts should state their sources, and, if appropriate, link to them. This matters to creators, because they deserve credit for their work. It also matters to readers who need to check the credibility of the original sources.

Unfortunately, some people on the web are not good internet citizens. This is particularly true with material that spreads quickly as hundreds or thousands of people share it – so-called “viral” content.

When that type of information travels around a network, people often fail to link it to sources, or they hide the sources altogether. For example, it was claimed that two million bikers were going to show up for President-elect Trump’s inauguration. Whatever your political persuasion, that would be pretty amazing.

But the source of the information, Right Alerts Polls, is not linked.
Here is a tip for locating the source. Using the Chrome web browser, select the text “Right Alerts Polls.” Then right-click your mouse (control-click on a Mac), and choose the option to search Google for the highlighted phrase.

Your computer will execute a search for “Right Alerts Polls.” To find the story, add “bikers” to the end of the search:
Our article appears at the top of the results list. Another option is to go directly to Google and type in the search terms.

So have we found the source? Not yet. When we click through to the supposed source article, we find that this article does not tell us where the information is coming from; however, it does have an extended quote from one of the “Two Million Bikers” organizers:

From the 2 Million Bikers to DC

“Those libtards need to be shut the hell up. This is not only a biker event but it is a Trump Supporters event. We are many and varied but we unite as one. Christian, Veterans, Bikers, Truckers, Steelworkers, Coal worker’s, Preachers, Fast food workers, all Colors on the human race, if you are a working class American and you support Mr. President Donald Trump then we would love to see you there. Our main objective is to show our Love for our new President and unite as one in a collective effort in Solidarity Brotherhood and Unity. God Bless America.”

So we just repeat our technique here, and select a bit of text from the quote and right-click/control-click. Our goal is to figure out where this quote came from, and searching on this small but unique piece of it should bring it close to the top of the Google results list.
When we search this snippet of the quote, we see that there are dozens of articles covering this story, using the same quote and sometimes even the same headline. But one of those results is the actual Facebook page for the event, and if we want a sense of how many people are committing, then this is a place to start.
As we begin to develop fact-checking habits, we should remember to not only scan the titles in a result list but also the URLs beneath the titles. These are clues as to which sources are best.

So we go to the “Two Million Biker” Facebook event page, and take a look. How close are they to getting two million bikers to commit to this?

So far, about 1,800 bikers have committed to going to the inauguration. That is quite positive – organizing is hard, and people have lives. Getting people to give up time for a political activity is tough. But it is well short of the “two million bikers” most of these articles were telling us were going to show up.

It appears that the overall numbers are “Mostly False” or “Unlikely” – there are people planning to attend, but the importance of the story was based around the scale of attendance. All indications seem to be that attendance will likely be a tenth of one percent (0.1%) of what the other articles promised. We only learned this by finding the source of the content.
Another type of content that often goes “viral” is photography. It is also some of the most difficult to track to a source. Note the example (below).

So, what is the story here? To get more information, insert the textual information off the image and do a Google search:
A *YouTube* video indicates that this was taken “outside a Portland, Oregon Walmart” and has been shared “hundreds of times since yesterday.” We will continue to trace the source with this new information. The next result shows why it is important to continue the search beyond the first set of results.

Which one of these items should I click on? Again, the idea here is to get closer to the actual event. One way to do that is to find the earliest post, but another way is to get geographically closer to the source. Who is more likely to get the facts of a local story correct, the local newspaper or a random blog?

As you scan the search results, look at the URLs. *Fox 13 News* has it in “trending.” *AmericaNow* has it in the
“society” section. But the *WGME* link has the story in a “news/local/” directory. This is interesting, because the other site said it happened in Oregon, and here the location is clearly Maine. This URL pattern is a strong point in the website’s favor.

Further indications that it might be a good source is that in the blurb it mentions the name of the photographer, “Matthew Mills.” The URL plus the specificity of the information tell us that we are on the right track.

This takes us to what looks like the news page where it went viral, which embeds the original post.
BIDDEFORD (WGME) -- A photo of a car taking up two parking spots surrounded by shopping carts at the Biddeford Walmart has been shared all over the country.

Matthew Mills of Old Orchard Beach took the picture of the parking lot justice Sunday afternoon.

CBS 13 posted it Monday morning, and it went viral.

OOPS! Someone wasn't happy about this double-parked car...Matthew Mills snapped this picture at the Biddeford Walmart.
We see that the first news report contained many inaccuracies. The photograph was not taken in Portland, Oregon—it was in Biddeford, which is near Portland, Maine. It has not been shared “hundreds of times” – it has been shared hundreds of thousands of times. And it was made viral by a CBS affiliate, a fact that ABC Action News in Tampa does not mention at all.

Now let’s look at the Facebook page where Matthew Mills shared it. We are looking to determine whether or not this went viral before CBS picked it up. We will double check that Mills is really from the Biddeford area and whether he was responsible for the shopping carts or just happened upon the scene.

The news post does not link back to the original, so we search for Matthew Mills again. We find some news outlets mentioning the original caption by Mills – “This guy got a lesson in parking.”

That is not the same as the caption that the news station put up – maybe it is what Mills originally used. We type “got a lesson in parking’ Matthew Mills” into Facebook, and we arrive at the original post.
Now we see an unfortunate practice perpetrated by some news organizations. They will cut other news organizations out of a story. In this example, *ABC Action News* said that the photograph had been shared hundreds of times, when in fact it was shared hundreds of thousands of times. If they had been completely truthful, *ABC Action News* would have mentioned that it was actually popularized by a *CBS* affiliate.

This practice can make it easier to trace something to the source. News organizations work hard to find the original source if it means they can cut other news organizations out of the picture. But it also tends to distort how virality happens. The picture here did not magically become viral—it became viral largely due to the reach of *WGME*.

Incidentally, we also find answers to other questions in the Matthew Mills version – he took the picture but didn’t arrange the carts, and he really is from Old Orchard Beach.
An extra verification step is to use Google Images to determine whether the photograph is a recycled image. Sometimes people take old images and pretend the images are their own – changing only the date and location. A Google reverse image search (see below) indicates that this does not appear to be the case in this instance; however, this is a very common type of viral photo called a “parking revenge” photo. The specific technique of circling carts around a double-parked car dates back to at least 2012.

When we click through we can see that the practice was popularized, at least to some extent, by Reddit users. Note the post from December 2012:
It is part of a parking revenge meme that dates back at least four years, and was popularized by Reddit. This particular photograph was shot by Matthew Mills in Biddeford, Maine. It became viral through the re-share provided by a local Maine news station.
13.

**USING GOOGLE REVERSE IMAGE SEARCH**

Most of the time finding the origin of an image on *Twitter* is easy. Just follow the links. For instance, take the chart in this tweet from *Twitter* user @NinjaEconomics. Should you evaluate it by finding out who @NinjaEconomics is?

No. Just follow the link to the source. Links are usually located in the last part of a tweet. As you can see, the chart is accompanied by more information about the data and how it was produced. It is from the Atlanta Federal Reserve, and it is the Federal Reserve, not @NinjaEconomics that you will evaluate.

The photo (below), was posted without a source and this generates questions.
First, is this actually a National Geographic photographer? More importantly, is this real? Is that lava so hot that it will literally set a metal tripod on fire? We are not lava experts so we will have to find the truth.

There is no link provided, so we will use the reverse image search. If you’re using Google Chrome as a browser, put the cursor over the photo and right-click (control-click on a Mac). A “context menu” will pop up and one of the options will be “Search Google for image.” For the sake of simplicity we will show solutions in this text as they would be implemented in Chrome.
When we reverse search this image, we find several pages that contain the photo from a variety of sites. One of the sites returned is Reddit. Reddit is a site that is famous for sharing these sorts of photos, but it also has a reputation for having a user base that is very good at spotting fake photos.
When we go to the Reddit page, we find there is an argument over whether the photo is fake or not. But again, Reddit is not our source here and we need to explore further. We click the link in the Reddit forum that says it is real, and get taken to an article where they actually talk to the photographer.
That brings us to one of the original stories about this photo:
We could stop here and just read the headline but all good fact-checkers know that headlines are not always truthful. So we read the article down to the bottom:

For this particular shot, Singson says, “Always trying to be creative, I thought it would be pretty cool (hot!) to take a lava pic with my shoes and tripod on fire while photographing lava.”

This may be a bit pedantic, but we still do not know if this was a staged event. Contrary to the headline, the photographer does not say lava made his shoes catch on fire. He says he wanted to take a picture of himself with his shoes on fire while standing on lava.

So did his shoes catch on fire, or did he set them on fire? At the bottom of this page, we can see that this is a retelling of an article published elsewhere; it is not this publication who talked to the photographer. It is a similar situation to what we saw in an earlier chapter, where the Blaze was simply retelling a story that was investigated by the Daily Dot.
In webspeak, “via” means you learned of a story or photo from someone else. In other words, we still have not located the source. So we go to the PetaPixel site from whence this photograph came. There we find an addendum on the original article:

Update on 7/17/13: Hawaii News Now confirms that this was done as a stunt:

The photo is real, but the flames are not the result of spontaneous combustion. “The flames on the tripod and my shoes did not start because of the lava” Singson freely admits. “It’s like if you put your shoe in a hot frying pan, it will not catch fire right away”.

Singson used an accelerant to start the flames then had his buddy snap the shot. The reason? “It’s just something I wanted for my Facebook cover photo”.

So a local news outfit has confirmed the photographer did use an accelerant. The photograph was indeed staged. Are we done now?

Not quite. You know what the next step is, right? Go to Hawaii News Now!

We click the link, and we find that the quote is good. Also, Hawaii News Now is a local news service, so they know quite a bit about lava fields. That is probably why they asked the question no one else seemed to ask – “Is that really possible?”

Finally, let’s find out about Hawaii News Now. We start by selecting Hawaii News Now and using our Google search option:
What we get back is pretty promising. There is a Google knowledge panel that comes up and tells us it is a bona fide local news program from a CBS affiliate in Hawaii.
We could stop here. We have solved this riddle. The photographer was really on hot lava, which is impressive in itself, but used some accelerant (such as lighter fluid) to set his shoes and tripod on fire. Additionally, the photo was a stunt, and not part of any naturally occurring *National Geographic* shoot. We’ve traced the story back to its source, found the answer, and got confirmation on the authoritative nature of the source.

We are sticklers for making absolutely sure of this; therefore, we are going to continue on, and click on the *Wikipedia* link to the article on the *Google* knowledge panel to make sure we are not missing anything.
FILTERING BY TIME AND PLACE TO FIND THE ORIGINAL

Finding the source of content is often a journey through time and space. The original story is also the first story, and as we saw with the Hawaiian news site, local sources often have special insights into stories.

There are specific functions you can use with Google and other search engines to help you find original material more quickly.

The following is an image that Twitter users have identified as another National Geographic photograph. But is it?
A Google reverse image search finds the photo, suggesting that the best search term is “birds attacking people.”

This suggestion is based on the fact that the pages where this photo appears often contain these words: “birds attacking people.”
We can modify that search, and we will retrieve only the older pictures.

We do that by clicking the “Tools” button and then using the “Time” drop-down to select “Custom range.” This should filter out some of the posts that merely include this in slideshows.
We pick a date in the past to see if we can filter out the newer photos. We remove the “birds attacking people” search and replace it with “bird,” since the other phrase sounds like a title for a slideshow with many of these sorts of photos in it. The original is not likely to be on a page like that; the slideshows come later in the viral cycle:
Why choose the year 2009? For viral photos, 2009 or 2010 is usually a good starting point. If you don’t find any results within that parameter, then go later, to a year like 2012. If you find too many results, then change the search to something like 2007.

Here we get a much better set of results. Instead of a list of “When Birds Attack” slideshows, we get a set of results talking about this specific photo. One of the results stands out.

This third result looks most promising for two reasons:

- The poster of the “Got too close to the hawk” result seems to know a bit more about the situation, noting “these birds are trained.”
- It mentions “Kazakhstan Eagle.” That is a name of a type of bird, but it’s also a place, and if we could confirm this incident took place in Kazakhstan, there will be other ways to trace this back to the original. Remember that our goal is about getting closer in time to the original, but it can also mean getting closer
geographically.

When we go to that page, it links us in the comments to a page that has the set of shots that the photographer was taking, as well as a shot of this cameraman being attacked from another angle.

It’s a series of photos from a hunting competition in Chengelsy Gorge, Kazakhstan. The eagle attacking him is tame and trained, but for some reason attacked him anyway. So this is real; it’s not photo-shopped or staged. At the same time it’s not a *National Geographic* photographer.
While this process takes some time to explain, in practice it can be done in about 90 seconds. Here’s a YouTube video that shows what this looks like in practice.

Note that as long as you are careful with confirmation bias, you can replace the search term “bird” with a term like “fake” to find pages claiming the image is fake and see what evidence they present.
ACTIVITY: TRACE VIRAL PHOTOS

The following two photographs have been attributed to *National Geographic* shoots by the same tweeter mentioned in the previous section.

If you are reading this on the web, go to it. If you are reading this book in PDF form, you’ll have to go find them at the Hapgood blog to use your *Google* reverse image search right-click/control-click action.

Bearing It

The first one is easy. Is this real, or fake? And are these *National Geographic* photographers or not? Is the bear real?
Swan Song

This second photograph is a lot harder. But is this real or fake? If real, can you find the name of the photographer in the swan as well as his nationality? If fake, can you find a debunking of it?

Truck Bomb

This next photograph is political. It was shared by a Twitter user who claimed it was a picture of an Irish Republican Army bombing. To paraphrase the poster – “This is London in 1993 after an IRA truck bomb. We didn’t ban Irish people or Catholics.” The poster is making a reference to recent moves to ban travel from Muslim countries to the United States.

Is this a picture of a 1993 London truck bombing? If so, how many people died and/or were injured? What was the response?
Going Rambo
Is this photograph real or fake?
PART IV
CREDIBILITY OF THE SOURCE
WHAT “READING LATERALLY” MEANS

Our third strategy entails fact-checkers reading across many connected sites instead of digging deep into the site at hand.

When you start to read a book, a journal article, or a physical newspaper you usually already know quite a bit about your source. You subscribed to the newspaper, or picked it up from a newsstand. You’ve ordered the book from Amazon or purchased it from a local bookstore because it was a book you were interested in reading. You’ve chosen a journal article either because of the quality of the journal article or because someone whose expertise and background you know cited it. In other words, when you get to the document you need to evaluate, the process of getting there has already given you some initial bearings.

Compared to these intellectual journeys, web reading is a bit more like teleportation. Even after tracing a source, you arrive at a page, site, and author that are often all unknown to you. How do you analyze the author’s qualifications or the trustworthiness of the site?

Researchers have found that most people go about this the wrong way. When confronted with a new site, they poke around the site and try to find out what the site says about itself by going to the “about page,” clicking around in onsite author biographies, or scrolling up and down the page. This is a faulty strategy for two reasons. First, if the site is untrustworthy, then what the site says about itself is most likely untrustworthy, as well. And, even if the site is generally trustworthy, it is inclined to paint the most favorable picture of its expertise and credibility possible.

The solution to this is, in the words of Sam Wineburg’s Stanford research team, to “read laterally.” Lateral readers don’t spend time on the page or site until they’ve first gotten their bearings by looking at what other sites and resources say about the source they are evaluating.

For example, when presented with a new site that needs to be evaluated, professional fact-checkers don’t spend much time on the site itself. Instead they get off the page and see what other authoritative sources have said about the site. They open up many tabs in their browser, piecing together different bits of information from across the web to get a better picture of the site they’re investigating. Many of the questions they ask are the same as the vertical readers scrolling up and down the pages of the source they are evaluating. But unlike those readers, they realize that the truth is more likely to be found in the network of links to (and commentaries about) the site than in the site itself.

Only when they’ve gotten their bearings from the rest of the network do they re-engage with the content. Readers gain a better understanding as to whether to trust the facts and analysis presented to them.
Lateral reading helps the reader understand both the perspective from which the site’s analyses come and if the site has an editorial process or expert reputation that would allow one to accept the truth of a site’s facts.

We’re going to deal with the latter issue of factual reliability, while noting that lateral reading is just as important for the first issue.
EVALUATING A WEBSITE OR PUBLICATION’S AUTHORITY

Authority and reliability are tricky to evaluate. Whether we admit it or not, most of us would like to ascribe authority to sites and authors who support our conclusions and deny authority to publications that disagree with our worldview. To us, this seems natural. The trustworthy publications are the ones saying things that are correct, and we define “correct” as what we believe to be true. A moment’s reflection will show the flaw in this way of thinking.

How do we get beyond our own myopia here? We can use Wikipedia’s guidelines for determining the reliability of publications. These guidelines were developed to help people with diametrically opposed positions argue in rational ways about the reliability of sources using common criteria.

For Wikipedians, reliable sources are defined by process, aim, and expertise. I think these criteria are worth thinking about as you fact-check.

Process

Above all, a reliable source for facts should have a process in place for encouraging accuracy, verifying facts, and correcting mistakes. Note that reputation and process might be apart from issues of bias. The New York Times is thought by many to have a center-left bias, the Wall Street Journal a center-right bias, and USA Today a centrist bias. Yet fact-checkers of all political stripes are happy to be able to track a fact down to one of these publications since they have reputations for a high degree of accuracy, and issue corrections when they get facts wrong.

The same thing applies to peer-reviewed publications. While there is much debate about the inherent flaws of peer review, peer review does get many eyes on data and results. Their process helps to keep many obviously flawed results out of publication. If a peer-reviewed journal has a large following of experts, that provides even more eyes on the article, and more chances to spot flaws. Since one’s reputation for research is on the line in front of one’s peers, it also provides incentives to be precise in claims and careful in analysis in a way that other forms of communication might not.
Expertise

According to Wikipedians, researchers and certain classes of professionals have expertise, and their usefulness is defined by that expertise. For example, we would expect a marine biologist to have a more informed opinion about the impact of global warming on marine life than the average person, particularly if they have done research in that area. Professional knowledge matters too. We would expect a health inspector to have a reasonably good knowledge of health code violations, even if they are not a scholar of the area. And while we often think researchers are more knowledgeable than professionals, this is not always the case. For a range of issues, professionals in a given area might have better insight than researchers, especially where a question deals with common practice.

Reporters, on the other hand, often have no domain expertise, but may write for papers that accurately summarize and convey the views of experts, professionals, and event participants. As reporters write in a niche area over many years (e.g. opioid drug policy) they may acquire expertise themselves.

Aim

Aim is defined by what the publication, author, or media source is attempting to accomplish. Aims are complex. Respected scientific journals, for example, aim for prestige within the scientific community, but must also have a business model. A site like the New York Times relies on ad revenue but is also dependent on maintaining a reputation for accuracy.

One way to think about aim is to ask what incentives an article or author has to get things right. An opinion column that gets a fact or two wrong won’t cause its author much trouble, whereas an article in a newspaper that gets facts wrong may damage the reputation of the reporter. On the far ends of the spectrum, a single bad or retracted article by a scientist can ruin a career, whereas an advocacy blog site can twist facts daily with no consequences.

Policy think tanks, such as the Cato Institute and the Center for American Progress, are interesting hybrid cases. To maintain their funding, they must continue to promote aims that have a particular bias. At the same time, their prestige (at least for the better known ones) depends on them promoting these aims while maintaining some level of honesty.

In general, you want to choose a publication that has strong incentives to get things right, as shown by both authorial intent and business model, reputational incentives, and history.
What are some quick techniques to identify an unfamiliar site’s worldview, process, aims, and expertise?

Web Searching a Domain

The simplest and quickest way to get a sense of where a site sits in the network ecosystem is to *Google* search the site. Since we want to find out what other sites are saying about the site while excluding what the site says about itself, we use a special search syntax that excludes pages from the target site.

For example, say we are looking at an article in the *Baltimore Gazette*:
The first presidential debate was held and Hillary Clinton was proclaimed the winner by the media. Indeed Clinton was able to turn in a strong debate performance, but did she do so fairly? Multiple reports and leaked information from inside the Clinton camp claim that the Clinton campaign was given the entire set of debate questions an entire week before the actual debate.
Is this a reputable newspaper?

The site is down right now, but when it was up, a search for “baltimoregazette.com” would have returned many pages, mostly from the site itself. As noted earlier, if we don’t know whether to trust a site, it doesn’t make much sense to trust the story the site tells us about itself.

So we use a search syntax that looks for all references to the site that are not on the site itself:

baltimoregazette.com -site:baltimoregazette.com

When we do that we get a set of results that we can scan, looking for sites we trust:

![Google search results](image)

These results, as we scan them, give us reason to suspect the site. Maybe we don’t know “City Paper,” which claims the site is fake. But we do know Snopes. When we take a look there, we find the following sentence about the Gazette:

On 21 September 2016, the Baltimore Gazette — a purveyor of fake news, not a real news outlet — published an article reporting that any “rioters” caught looting in Charlotte would permanently lose food stamps and all other government benefits...

From Snopes, that’s pretty definitive. This is a fake news site.

Searches like this don’t always turn up Snopes or Politifact. Here’s the site of the Pacific Justice Institute:
Here, a search of Google turns up a Wikipedia article:

Figure 53

That article explains that this is a conservative legal defense fund that has been named a hate site by the Southern Poverty Law Center.

Maybe to you that means that nothing from this site is trustworthy; maybe to another person it simply means proceed with caution. But after a short search and two clicks, you can begin reading an article from this site with a better idea of the purpose behind it, a key ingredient of intentional reading.
Finding Out Who Runs a Site with WHOIS and Other Tools

Some smaller sites don’t have reliable commentary around them. For these sites, using WHOIS to find who owns them may be a useful strategy.

WHOIS gets you information about who is the administrator of the site domain. It can be done from your computer’s command line in many cases, but here we’ll show the ICANN interface, where we are searching to see who owns Mother Jones, an online news site:

![ICANN WHOIS](image)

**Showing results for: MOTHERJONES.COM**

Original Query: motherjones.com

### Contact Information

<table>
<thead>
<tr>
<th>Registrant Contact</th>
<th>Admin Contact</th>
<th>Tech Contact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name: Foundation for National Progress</td>
<td>Name: Foundation for National Progress</td>
<td>Name: Wise, Robert</td>
</tr>
<tr>
<td>Mailing Address: 222 Sutter Street, San Francisco CA 94108 US</td>
<td>Mailing Address: 222 Sutter Street, San Francisco CA 94108 US</td>
<td>Mailing Address: 222 Sutter Street, San Francisco CA 94108 US</td>
</tr>
<tr>
<td>Phone: +1.4153211700</td>
<td>Phone: +1.4153211700</td>
<td>Phone: +1.4153211700</td>
</tr>
<tr>
<td>Ext:</td>
<td>Ext:</td>
<td>Ext:</td>
</tr>
<tr>
<td>Fax: +1.4153211701</td>
<td>Fax: +1.4153211701</td>
<td>Fax: +1.4153211701</td>
</tr>
<tr>
<td>Fax Ext:</td>
<td>Fax Ext:</td>
<td>Fax Ext:</td>
</tr>
<tr>
<td>Email: <a href="mailto:rwise@motherjones.com">rwise@motherjones.com</a></td>
<td>Email: <a href="mailto:rwise@motherjones.com">rwise@motherjones.com</a></td>
<td>Email: <a href="mailto:rwise@motherjones.com">rwise@motherjones.com</a></td>
</tr>
</tbody>
</table>

Figure 55

When we search on the owner, we find that:

The Foundation for National Progress is a nonprofit organization created to educate the American public by publishing Mother Jones. Mother Jones is a multiplatform news organization that conducts in-depth investigative reporting and high quality, original, explanatory journalism on major social issues, including money in politics, gun violence, economic inequality and the future of work.
(We could have found this out by other means as well, of course).

Unfortunately, WHOIS blockers have dramatically reduced the value of WHOIS searches. The famous *Baltimore Gazette* fake news site from 2016, for example, uses a proxy service to hide revealing information.

![ICANN WHOIS](image)

**Showing results for: BALTIMOREGAZETTE.COM**
Original Query: baltimoregazette.com

<table>
<thead>
<tr>
<th>Contact Information</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Registrant Contact</strong></td>
</tr>
<tr>
<td>Name: Registration Private</td>
</tr>
<tr>
<td>Organization: Domains By Proxy, LLC</td>
</tr>
<tr>
<td>Mailing Address: DomainsByProxy.com, Scottsdale AZ 85260 US</td>
</tr>
<tr>
<td>Phone: +1.4806242599</td>
</tr>
<tr>
<td>Ext:</td>
</tr>
<tr>
<td>Fax: +1.4806242598</td>
</tr>
<tr>
<td>Fax Ext:</td>
</tr>
<tr>
<td>Email: BALTIMOREGAZETTE.COm</td>
</tr>
<tr>
<td><a href="mailto:M@domainsbyproxy.com">M@domainsbyproxy.com</a></td>
</tr>
</tbody>
</table>

Figure 56

The owner of the site here isn’t Domains by Proxy, as the record indicates. Instead, Domains by Proxy is a service, often available for a couple dollars a year, that obscures the true ownership of the site. These masking services are starting to become the norm, dramatically reducing the usefulness of WHOIS searches.

That said, there is still useful information to be had here, particularly in the date the baltimoregazette.com domain was registered, which is listed here as being mid-2015:
Important Dates

Updated Date: 2016-07-18
Created Date: 2015-07-17
Registration Expiration Date: 2017-07-17

Figure 57
If this were an established local paper, it would be fairly odd for it to have first registered the site a year ago.
19.

ACTIVITY: EVALUATE A SITE

Evaluate the reputations of the following sites by “reading laterally.” Answer the following questions to determine the reputability of each site: Who runs them? To what purpose? What is their history of accuracy, and how do they rate on process, aim, and expertise?

4. http://www.nature.com/nature/journal/
8. http://occupydemocrats.com/2017/02/11...
There’s no more dreaded phrase to the fact-checker than “a recent study says.” Recent studies say that chocolate cures cancer, prevents cancer, and may have no impact on cancer. Recent studies say that holding a pencil in your teeth makes you happier. Recent studies say that the scientific process is failing, and others say it is just fine.

Most studies are data points – emerging evidence that lends weight to one conclusion or another but does not resolve questions definitively. What we want as a fact-checker is not data points, but the broad consensus of experts. And the broad consensus of experts is rare.

The following chapters are not meant to show you how to meticulously evaluate research claims. Instead, they are meant to give you, the reader, some quick and time-saving ways to decide what sorts of research can be safely passed over when you are looking for a reliable source. We take as our premise that information is abundant and time is scarce. As such, it’s better to err on the side of moving onto the next article than to invest time in an article that displays warning signs regarding either expertise or accuracy.
FINDING A JOURNAL’S IMPACT FACTOR

As mentioned earlier, the process of evaluating web content is one of elimination. In a world where information is plentiful, we can be a bit demanding about what counts as evidence. When it comes to research, one would expect that published academic research cited for a claim comes from respected peer-reviewed journals.

Consider this journal:

![Journal Image]

Figure 58

We ask ourselves, “Is it a journal that gives authority to this article?” Or “Is it just another web-based paper mill?”

**Impact Factor**

Our first check is to find the journal’s “impact factor” (IF). The impact factor is a measure of the frequency
with which the average article in a journal has been cited in a particular year. It is used to measure the importance or rank of a journal by calculating the times its articles are cited.

**How Impact Factor is Calculated?**

The calculation is based on a two-year period, and involves dividing the number of times articles were cited by the number of articles that are citable.

While a flawed metric for assessing the relative importance of journals, impact factor is a useful tool for quickly identifying journals that are not part of a known circle of academic discourse, or that are not peer-reviewed.

When we search Google for PLOS Medicine impact factor, we retrieve a knowledge panel with an impact factor of 13.585 (see below).

---

Figure 59

Impact factors can go into the 30s, but we’re using this as a quick elimination test, not a ranking, so we’re happy with anything over 1. We still have work to do on this article, but it’s worth keeping in the mix.

What about this one?
When we look for the impact factor for the *Journal of Obesity and Weight-loss Medication*, we get a result with a link to this journal at the top of the result list, but no panel, as there is no registered impact factor for this journal.

The article may be excellent but we do not know for sure. Likewise, there are occasionally articles published in the most prestigious journals that are pure junk. Be careful in your use of impact factor; a journal with an impact factor of 10 is not necessarily better than a journal with an impact factor of 3, especially if you are dealing with a niche subject.

But after a quick analysis, the *PLOS Medicine* article seems more trustworthy than the *Journal of Obesity and Weight-loss Medication* article. In fact, if you were deciding whether to re-share a story in your feed and the evidence for the story came from the *Obesity* journal, we would recommend skipping this source.
USING GOOGLE SCHOLAR TO CHECK AUTHOR EXPERTISE

Not all, or even most, expertise is academic. But when the expertise cited is academic, scholarly publications by the researcher can go a long way to establishing their position in the academic community.

Let’s look at David Bann, who wrote the *PLOS Medicine* article we looked at in the last section. To do that, we go to Google Scholar (not the general page) and type in his name (below).

![Google Scholar search results for David Bann](image)

The results reveal some interesting bits of information. First, Bann has a history of publishing in the area of lifespan obesity patterns. At the bottom of each result, we see how many times each article he authored or...
co-authored is cited. These are not amazing numbers, but for a niche area they are a healthy citation rate. Many articles published are not cited at all, but at least one of his works has over 100 citations.

Additionally, if we scan down the right side of the page, we see some names we might recognize – the National Institutes of Health (NIH) and another PLOS article.

Remember that we are looking for expertise in the area of the claim. These are great credentials on the topic of obesity. They would not be great if we were looking for evidence of Bann’s expertise on the topic of opiate addiction. Happily, for this exercise, we are only concerned about the topic of obesity.

By point of comparison, we can look at a publication in Europhysics News that attacks the standard view of the 9/11 World Trade Center collapse. We see this view represented in this story on a popular alternative news and conspiracy site called AnonHQ:

Figure 63

The journal cited is Europhysics News, and when we look it up in Google, we find no impact factor. In fact, a short investigation of the journal reveals it is not a peer-reviewed journal, but a magazine associated with the European Physics Society. The author is being untruthful, or does not understand the difference between a scientific journal and a scientific organization’s magazine.

But what about the authors? Do they have a variety of papers on the mathematical modeling of building demolitions? If you search for the names in Google Scholar, you will find that at least one of the authors does have some modelling experience on architectural stresses, although most of his published work was from years ago.
It is fair to say that this article was not peer-reviewed and should not be treated as a substantial contribution to the body of research on the 9/11 collapse. The headline of the blog article that brought us here is false, as is their claim that a *European Scientific Journal* concluded 9/11 was a controlled demolition.

But it is worthwhile to note that at least one of the authors writing this paper does have some expertise in a related field. We are left with the question – What does “generally” mean in the phrase “Experts generally agree on X.”

What should we do with this article? Well, it is an article published in a non-peer-reviewed journal by an expert who published a number of other respected articles (though quite a long time ago, in one case). To an expert, that definitely could be interesting. To a new fact-checker looking for the majority and significant minority views of the field, it is probably not the best source.
How should we think about research articles? Often we think that newer is better, and sometimes this is true. For example, new phones are better than old phones; and new textbooks are more up-to-date than old textbooks. But the understanding many students have about scholarly articles is that newer studies “replace” older studies.

In general, that is not how science works. In science, multiple conflicting studies are published over long periods of time, and each one is a drop in the bucket of the claim it supports. Over time, the weight of the evidence ends up on one side or another. Depending on the quality of the new research, some drops are bigger than others (some much bigger), but overall it is an incremental process.

Studies that are consistent with previous research are often more trustworthy than those that have surprising or unexpected results. This runs counter to the narrative promoted by the press – “news,” after all, favors what is new and different. The unfortunate effect of the press’s presentation of science (especially in the field of health) is that they would rather not give a sense of the slow accumulation of evidence for each side of an issue. Their narrative often presents a world where last month’s findings are “overturned” by this month’s findings, which are then, “overturned” back to the original finding a month from now. This whiplash presentation, such as “Chocolate is good for you! Chocolate is bad for you!” undermines the public’s faith in science. But the whiplash is not the fault of science. It is a product of the inappropriate presentation by the press.

As a fact-checker, your job is not to resolve debates based on new evidence, but to accurately summarize the state of research and the consensus of experts in a given area, taking into account majority and significant minority views.

For this reason, fact-checking communities such as Wikipedia discourage authors from over-citing individual research, which tends to point in different directions. Instead, Wikipedia encourages users to find high quality secondary sources that reliably summarize the research base of a certain area, or research reviews of multiple works. This is good advice for fact-checkers as well. Without an expert’s background, it can be challenging to place new research in the context of old, which is what you want to do.

Below are two claims that ran recently in the Washington Post:

The alcohol industry and some government agencies continue to promote the idea that moderate drinking provides some health benefits. But new research is beginning to call even that long-standing claim into question.

Reading down further, we find a more specific claim; the medical consensus is that alcohol is a carcinogen even at low levels of consumption. Is this true?
The first thing we would do is look at the authorship of the article. It is from the *Washington Post*, which tends to be a reliable publication, and one of its authors has made a career of data analysis (and won a Pulitzer prize as part of a team that analyzed data and discovered election fraud in a Florida mayoral race). Keep in mind that some writers may be better interpreters of the data than we are.

But suppose we want to dig further and find out if they are really looking at a shift in the expert consensus, or just adding more drops to the evidence bucket. How would we do that?

First, we would check to find out where the pieces were published. The *Post* article mentions two articles by “Jennie Connor, a professor at the University of Otago Dunedin School of Medicine,” one published in 2016 and the other published earlier. Let’s find the more recent one, which seems to be a key source for this article. We go to *Google Scholar* and type in “Jennie Connor 2016”.

![Google Scholar search results](image)

Figure 65

We scan the results and decide that the top result is probably what we are looking for; however, we also notice that Connor has written other well-cited articles in the field of health.

What about her article on “Alcohol consumption as a cause of cancer”? It was published in 2017 (which is probably the physical journal’s publication date, the article having been released in 2016). Nevertheless, it has already been cited by twelve other papers.

What about this publication *Addiction*? Is it reputable? We will do an impact factor search.
Figure 66

Yes it looks legitimate. We also note in the knowledge panel to the right that the journal was founded in the 1880s. If we click through to that Wikipedia article, it tells us that this journal ranks second in impact factor for journals on substance abuse.

Again, you should never use impact factor for fine-grained distinctions. What we are verifying is that the Washington Post wasn’t fooled into covering some research far out of the mainstream of substance abuse studies, or tricked into covering something published in a sketchy journal. It is clear from this quick check that this is a researcher well within the mainstream of her profession, publishing in prominent journals.

Next we ask what kind of article this is. Sometimes journals publish short reactions to other works, or smaller opinion pieces. What we’d like to see here is that this was either new research or a substantial review of research. We find from the abstract that it is primarily a review of research, including some of the newer studies. We note that it is a six-page article, and therefore not likely to be a simple letter or response to another article. The abstract also goes into detail about the breadth of evidence reviewed.

Frustratingly, we cannot get our hands on the article, but this probably tells us enough about it for our purposes.
Let’s continue with the “alcohol is closely associated with cancer” claim from the last section. Let’s see if we can get a decent summary from a respected organization that deals with these issues.

This takes a bit of domain knowledge, but for information on disease, the United States’ National Institutes of Health (NIH) is considered one of the leading authorities. What do they say about this issue?

Alcohol Consumption and the Risk of Cancer
https://pubs.niaaa.nih.gov/Pubs/AlcoholAndCancerRisk.pdf
by A Policy • Related articles
Alcohol consumption has been linked to an increased risk for various types of cancer. A combined analysis of more than 200 studies assessing the link between ...

Scientists rethink alcohol/breast cancer relationship | National Institute ...
https://www.niaaa.nih.gov/pubs/scientists-rethink-alcohol-breast-cancer-risk...
Oct 16, 2012 • Time course and metabolism are important factors. According to National Institute on Alcohol Abuse and Alcoholism (NIAAA) scientists, existing ...

Alcohol Flush Signals Increased Cancer Risk among East Asians ...
Mar 23, 2009 • National Institutes of Health (NIH) • Turning Discovery into Health ... Alcohol Flush Signals Increased Cancer Risk among East Asians ...

Alcohol Consumption and Cancer Risk - NCBI Bookshelf
by O Oyesanmi - 2010 • Cited by 27 • Related articles
The purpose of this report is to systematically examine the possible causal mechanism(s) that may explain the association between alcohol (ethanol) ...

Alcohol and Cancer Risk Fact Sheet - National Cancer Institute
A fact sheet that summarizes the evidence linking alcohol consumption to the risk of various cancers. Includes information about factors that affect the risk of ...
What we don’t want is a random article. We are not experts and we don’t want to have to guess at the weights to give individual research. We are looking for a summary.

As we scan the results we find a “risk fact-sheet” from the National Cancer Institute that has a “gov” domain. Unlike some domain suffixes (i.e. com, org, net) a gov domain belongs to the federal government of the United States and these sites are strictly regulated. Also, a fact sheet is a summary, so we click through.

Based on extensive reviews of research studies, there is a strong scientific consensus of an association between alcohol drinking and several types of cancer (1, 2). In its Report on Carcinogens, the National Toxicology Program of the US Department of Health and Human Services lists consumption of alcoholic beverages as a known human carcinogen. The research evidence indicates that the more alcohol a person drinks—particularly the more alcohol a person drinks regularly over time—the higher his or her risk of developing an alcohol-associated cancer. Based on data from 2009, an estimated 3.5 percent of all cancer deaths in the United States (about 19,500 deaths) were alcohol related (3).

With the “.gov” extension, this page is very likely to be linked to the NIH. But just in case, we Google search the site to see who runs it and what their reputation is (see below).
National Cancer Institute. We are simply checking our impression that this is an authoritative body of the NIH. Note the blurb from the fifth result down:

The National Cancer Institute (NCI) is part of the National Institutes of Health (NIH), which is one of 11 agencies that compose the Department of Health and Human Services (HHS). The NCI, established under the National Cancer Institute Act of 1937, is the Federal Government’s principal agency for cancer research and training.

As always, we glance up to the web address and make sure we are really getting this information from the NIH. If we were a researcher, we would sort through more of this. We might review individual articles or make sure that some more out-of-the-mainstream views are not being ignored. Such an effort would take a deep background and understanding of the underlying issues. But we’re not researchers. We’re just people looking to find out if our rationalization for those two after-work drinks is bogus. And on that level, it’s not looking particularly good for us. We have found a major review of the evidence in a major journal stating there is really no safe level of drinking when it comes to cancer. We also have the NIH, one of the most trusted sources of health information in the U.S. telling us in an FAQ that there is a strong consensus that alcohol consumption predicts cancer.
In the last section, we selected a very trustworthy source of medical consensus (the NIH) and read that alcohol consumption predicts cancer.

This is an important technique to have in your research mix. Too often, we execute web search after web search without first asking who would constitute as an expert. Unsurprisingly, when we do things in this order, we end up valuing the expertise of people who agree with us and devaluing the expertise of those who don’t. If you find yourself going down a rabbit hole of conflicting information in your searches, back up a second and ask yourself – whose expertise would you respect? Maybe it’s not the NIH. Maybe it’s the Mayo Clinic, or Medline, or the World Health Organization. But deciding who has expertise before you search, will mediate any tendencies toward confirmation bias.

Given the evidence we have seen in previous chapters about alcohol and cancer, will we give up our after-work glass of wine? The evidence is still emerging, and maybe the risk is worth it. But the Washington Post article is not the newest version of “eating grapefruit will make you thinner.” It is not even “Nutrasweet may make you fat,” which is an interesting finding, but a point around which there is no consensus. Instead “small amounts of daily alcohol increase cancer risk” represents a real emerging consensus in the research, and from our review we find it’s not even a particularly new trend. The consensus emerged some time ago (the NIH FAQ dates back to 2010) but has been poorly communicated to the public.
EVALUATING NEWS SOURCES

Evaluating news sources can be very contentious. People have their favorite news sources and do not like to be told that their news source is untrustworthy.

For fact-checking, it’s helpful to draw a distinction between two activities:

- **News gathering**: Where news organizations do investigative work – calling sources, researching public documents, and checking and publishing facts. For example, getting the facts of Bernie Sanders involvement in the passage of several bills.

- **News analysis**: Takes those facts and strings them into a larger narrative. For example, “Senator Sanders an effective legislator behind the scenes” or “Senator Sanders largely ineffective Senator behind the scenes.”

Most newspaper articles are not lists of facts, which means that publications like the *Wall Street Journal* and the *New York Times* do both news gathering and news analysis in stories. What has been lost in the dismissal of the *New York Times* as liberal and the *Wall Street Journal* as conservative is that these are primarily biases of the news analysis portion of what they do. To the extent the bias exists, it’s in what they choose to cover, to whom they choose to talk, and what they imply in the way they arrange those facts they collect.

The news gathering piece is affected by this, but in many ways largely separate, and the reputation for fact checking is largely separate as well. *MSNBC*, for example, has a liberal slant to its news, but a smart liberal would be more likely to trust a fact in the *Wall Street Journal* than a fact uttered on *MSNBC* because the *Wall Street Journal* has a reputation for fact-checking and accuracy that *MSNBC* does not. The same holds true for someone looking at the *New York Observer* vs. the *New York Times*. Even if you like the perspective of the *Observer*, if you were asked to bet on the accuracy of two pieces – one from the *Observer* and one from the *Times* – you could make a lot of money betting on the *Times*.

Narratives are a different matter. You may like the narrative of *MSNBC* or the *Observer* – or even find it more in line with reality. You might rely on them for insight. But if you are looking to validate a fact, the question you want to ask is not always “What is the bias of this publication?” but rather, “What is this publication’s record with concern to accuracy?”
WHAT MAKES A TRUSTWORTHY NEWS SOURCE?

Experts have looked extensively at what sorts of qualities in a news source tend to result in fair and accurate coverage. Sometimes, however, the number and complexity of the various qualities can be daunting. We suggest the following short list of things to consider.

- **Machinery of care**: Good news sources have significant processes and resources dedicated to promoting accuracy, and correcting error.
- **Transparency**: Good news sources clearly mark opinion columns as opinion, disclose conflicts of interest, indicate in stories where information was obtained and how it was verified, and provide links to sources.
- **Expertise**: Good news sources hire reporters with reporting or area expertise who have been educated in the processes of ethical journalism. Where new writers with other expertise are brought in, they are educated by the organization.
- **Agenda**: The primary mission of a good news source is to inform its readers, not elect Democrats, promote tax cuts, or reform schools. You should absolutely read writers with activist missions like these, but do not treat them as “pure” news sources.

Here’s an important tip – *approach agenda last*. It is easy to see bias in people you disagree with, and hard to see bias in people you agree with. But bias isn’t agenda. Bias is about how people see things; agenda is about what the news source is set up to do. A site that clearly marks opinion columns as opinion, employs dozens of fact-checkers, hires professional reporters, and takes care to be transparent about sources, methods, and conflicts of interest is less likely to be driven by political agenda than a site that does not do these things. And this holds even if the reporters themselves may have personal bias. Good process and news culture goes a long way to mitigating personal bias.

Yet, you may see some level of these things and still have doubt. If the first three indicators do not settle the question for you, you should consider agenda. Is the source connected to political party leadership? Funded by oil companies? Have the owners made comments about what they are trying to achieve with their publication, and are those ends about specific social or political change or about creating a more informed public?

We cannot stress enough that you should read things by people with political agendas. It is an important part of your news diet. It is also the case that sometimes the people with the most expertise, work for organizations
that are trying to accomplish social or political goals. But when sourcing a fact or a statistic, agenda can get in
the way and you would want to find a less agenda-driven source if possible.
When it comes down to accuracy, there are a number of national newspapers in most countries that are well-staffed with reporters and have an editorial process that places a premium on accuracy. These papers are sometimes referred to as “newspapers of record.”¹ We’re aware that the origin of the term was originally a marketing plan to distinguish the *New York Times* from its rivals. At the same time, it captures an aspiration that is not common across many publications in a country.

National newspapers of record are distinguished in two ways:

1. They are rigorous, showing attention to detail and having accountability in their editorial processes.
2. They have a truly national view and attempt to be the best possible record of what happened in the nation (not just a region) on a given day.

The United States is considered by some to have at least four national newspapers of record:

- *The New York Times*
- *The Wall Street Journal*
- *The Los Angeles Times*
- *The Washington Post*

You could add in the *Boston Globe*, *Miami Herald*, or *Chicago Tribune*. Or subtract the *LA Times* or *Washington Post*. These lists are meant to be starting points, indicating that a given publication has a greater reputation and reach than, say, the *Clinton Daily Item*.

Some other English-language newspapers of record:

- *The Times* (UK)
- *The Daily Telegraph* (UK)
- *The Irish Times* (Ireland)
- *The Times of India* (India)
- *New Zealand Herald* (New Zealand)
Does that mean these papers are the arbiters of truth? No. Where there are disagreements between these papers and other reputable sources, it could be worth investigating.

As an example, in the run up to the Iraq War, the Knight Ridder news agency was in general a far more reliable news source on issues of faulty intelligence than the New York Times. In fact, reporting from the New York Times back then was particularly bad, and many have pointed to one reporter in particular, Judith Miller, who was far too credulous in repeating information fed to her by war hawks. Had you relied on just the New York Times for your information on these issues, you would have been misinformed.

There is much to be said about failings such as this, and it is certainly the case that high profile failings such as these have eroded faith in the press more generally, and, for some, created the impression that there really is no difference between the New York Times, the Springfield Herald, and your neighbor’s political Facebook page. This is, to say the least, overcompensation. We rely on major papers to tell us the truth, and rely on them to allocate resources to investigate and present that truth with an accuracy hard to match on a smaller budget. When they fail, as we saw with Iraq, horrible things can happen. But that is as much a testament to how much we rely on these publications to inform our discourse as it is a statement on their reliability.

A literate fact-checker does not take what is said in newspapers of record as truth. But, likewise, any person who doesn’t recognize the New York Times or Sydney Morning Herald as more than your average newspaper is going to be less than efficient at evaluating information. Learn to recognize the major newspapers in countries whose news you follow to assess information more quickly.
ACTIVITY: EXPERT OR CRANK?

Twitter Expertise?

This guy has a pretty negative reaction to something published in a highly reputable journal. Is he an expert, or just a guy with opinions about things?

Bullshit! Aztec society collapsed in 1519 fr. Cortes & smallpox. Salmonella in 1540 was far too late. And the painting is European fantasy. https://t.co/FZUbbyPiOo
— Michael E. Smith (@MichaelESmith) February 17, 2017

Figure 69

Woodward and Bernstein

Are these the reporters who brought down Nixon? Is this a trustworthy reporter sharing this photo?

This is Woodward and Bernstein. Nixon called them the enemy. They proved that no president is above the law. #NotTheEnemy pic.twitter.com/mU5Cx3JCuH
— pixelatedboat aka “mr tweets” (@pixelatedboat) February 19, 2017

Figure 70
ACTIVITY: FIND TOP AUTHORITIES FOR A SUBJECT

Get together in small groups, and by both pooling group knowledge and doing research, develop a list of three authoritative books/websites for information on one of the following subjects:

- Lead Poisoning in Homes
- American Hate Groups
- Weight Loss
- Nuclear Policy
- Philosophy of Stoicism
- Civil War and Race
- The Deficit and Deficit Spending

Your sources should be:

- Written, edited, or curated by people with a high-level of subject expertise
- Have a process in place that ensures accuracy and verification

Or, the sources should be:

- Written and published by someone with a reputation for accuracy
- Informed by a broad array of expert interviews and perspectives

When each group has finished their selection, trade your list of expert books/sites with another group and have that group critique the list.

Some questions for reflection:

- Do the best sources show up at the top of Google search results?
- If not, what does show up? Why do you think that result shows up instead?
One relatively common form of misinformation is the fake celebrity retweet. Sometimes this happens by accident—a person mistakenly retweets a parody account as real. Sometimes this happens by design, with an account faking a retweet. Here are some tips to make sure that the tweet you are looking at on Twitter is from the person you are attributing it to.

**Twitter Identity Basics**

Twitter accounts are generally run by a single person. However, unlike Facebook, Twitter does not enforce a “real name” policy, which makes it easy for one person to run multiple accounts, and to run accounts under different names. In fact, an important part of Twitter culture is the constellation of parody accounts, bots, and single issue accounts that amuse and inform Twitter subscribers.

At the same time, it is easy to get confused. As an example, consider the account of Representative Jack Kimble. Here’s a typical tweet:

![Jack Kimble](https://example.com/jack-kimble-profile-picture)

**Jack Kimble**

@RepJackKimble

*Why have the wars cost so much under Obama? Check the budgets. Bush fought 2 wars without costing taxpayers a dime*

Figure 71

If you’re a liberal, looking at this tweet may get your blood boiling. How can anyone possibly believe this? Especially a Representative?

Scanning the Twitter bio doesn’t help.
Figure 72

Here we see that he’s from the 54th District of California and he has written a book. Now if we are reading carefully we might notice some fishy things here: his book, *Profiles in Courageousness*, seems like a parodic re-titling of Jack Kennedy’s *Profiles in Courage*. “E pluribus unum,” which means “From the many, one,” is translated to “1 nation under God”.

Also, California only has 53 districts.

Unfortunately, you may be so upset about the comments that you won’t notice any of these things. So what is a general purpose indicator that you need to slow down? In most cases, it’s going to be the absence of a “verified account” marker.

**Checking Verified Accounts**

As a counter-example to “Representative Kimble,” here’s a real representative, Jason Chaffetz, from Utah’s 3rd District.
That little blue seal with the check mark (the “verified badge”) indicates that this is a “verified identity” by Twitter – Twitter asserts that this person has proven that they are who they say they are.

Who gets to get verified? It’s a bit unclear. Twitter puts it this way:

An account may be verified if it is determined to be an account of public interest. Typically this includes accounts maintained by users in music, acting, fashion, government, politics, religion, journalism, media, sports, business, and other key interest areas.

However, all members of Congress and senior administration officials qualify for such status. So do most major public figures and prominent writers. If you don’t see the blue badge, either disregard the tweet as suspicious, or do further research.

One additional note: sometimes people try to fake these indicators; an example is faking a verification symbol in a header.

---

Figure 73

That little blue seal with the check mark (the “verified badge”) indicates that this is a “verified identity” by Twitter – Twitter asserts that this person has proven that they are who they say they are.

Who gets to get verified? It’s a bit unclear. Twitter puts it this way:

An account may be verified if it is determined to be an account of public interest. Typically this includes accounts maintained by users in music, acting, fashion, government, politics, religion, journalism, media, sports, business, and other key interest areas.

However, all members of Congress and senior administration officials qualify for such status. So do most major public figures and prominent writers. If you don’t see the blue badge, either disregard the tweet as suspicious, or do further research.

One additional note: sometimes people try to fake these indicators; an example is faking a verification symbol in a header.
Figure 74

This user has used their background image to place a verification badge next to their name. To steer clear of these sorts of hacks, always view the badge in the sidebar or small “hover” card, not in the header. To be extra sure it is legitimate, hover your cursor over it, and the words “verified account” should pop up.

This sounds complicated, but once you learn it, it only takes a few seconds. In the example (below) we use the quick hover technique to check to see if this is really New York Governor Andrew Cuomo’s Spotify playlist, or a fake account.
Other Methods

Not all celebrities have verified accounts. If you don’t find the verification badge, you may have to dig a little deeper.

There are a couple things to look for in an unverified account:

- **Start date**: Did the user fire up this account six weeks ago? In general, older accounts are more trustworthy.
- **Followers**: Not always a perfect metric, but do the number of followers seem about right for the personality’s popularity? Do they have any followers you know?
- **Previous Tweets**: Are there many previous tweets, and are they what you would expect from the account? Do they have conversations with people in ways that you would expect?
As an example, here is the Minerva Schools’ Twitter account. Minerva is a small, but high-profile school in California. The account is not verified. Is the account legitimate? Is it really Minerva?

Figure 76

A number of things suggest it is. It was created in August 2013, right around the time that Minerva was created. It has followers we recognize from educational technology, which is what the school is known for. One of the followers is a person that we know works at Minerva.
Figure 77

We could stop there, or we could also note that the tweet stream is entirely consistent with what we would expect for an organization like this, and the number of followers, while not huge, is in line with what we might expect for an account like this.

No one single factor clinches it (although the employee showing up in the follow list comes close), but all these factors together give us a fair amount of confidence that this is a legitimate account.

If we wanted to go one step further, we could web search the handle and see if it is referenced from any official pages.
Fake Screenshots

Sometimes people fake screenshots of tweets that never happened.

Not all tweet screenshots are fake. Many times Twitter users will screenshot a tweet rather than retweet it because they fear the original will be deleted. Below is Michael Li screen-shotting an embarrassing tweet which was later deleted.

![Tweet by Michael Li](image1)

When Dan Patrick has a twitter typo, he has one heck of a twitter typo.

![Tweet by Dan Patrick](image2)

MARRIAGE= ONE MAN & ONE MAN. Enough of these activist judges. FAVORITE if you agree. I know the silent majority out there is with us!

![Tweet by Dan Patrick](image3)

![Retweets: 1,037 Likes: 718](image4)

12:13 PM - 26 Feb 2014

91 1.0K 718

Figure 78

Other times, people may screenshot a tweet because they wish to discuss a tweet without attracting the ire of a particular group of followers. As an example, during the #Gamergate controversy many people critical of Gamergate took screenshots of bad behavior on Twitter (harassment and the like) because they were afraid that if they commented via re-tweeting they might become a target themselves.
Sometimes people retweet screenshots as a way of breaking a chain of credit, so that people will be forced to retweet them, and not the original tweeter. This practice is rightfully frowned on.

Sometimes the screenshot may be fabricated. In fact, many “tweet generators” exist online that allow you to create fake pictures of tweets. The tweet (below) is fabricated.

Figure 79

If you come across a person re-tweeting a screenshot, check to see if the tweet really exists on Twitter. In the above example, you could check Obama’s timeline.

**Deleted Tweets**

What if they deleted the tweet, as in the “ONE MAN + ONE MAN” example above? How do you verify it then? Or what if the tweet someone was referencing has since been deleted.

Don’t worry – in many cases there are still ways to find the tweet.

If it’s a tweet from a politician (and it usually is) you can try Politiiwhoops, which logs all tweets deleted by significant public officials. Here are some tweets recently deleted by President Trump:
Figure 80

Another technique is searching for the Twitter account in Google and looking for the cached version of the page. In the video below, we search for @RealDonaldTrump in Google and then look at the cached version of his Twitter page. This works well with things recent enough to be on the first page of a Twitter stream, but old enough that Google has indexed them.
Figure 81

The Twitter bar sometimes obscures the cache information, but if you can see it, it will tell you when it was last indexed. The time is in Greenwich Mean Time (the same time as London, England). So for instance, this cache of Trump's tweets was taken at 2 o'clock London time.

Figure 82
ACTIVITY: VERIFY A TWITTER ACCOUNT

Kellogg’s Rant

General Kellogg was promoted by President Trump as acting head of the National Security Council on February 13, 2017. Is this Twitter account his?

https://twitter.com/GenKeithKellogg/status/832825494009638912

Explain your reasons.
Sometimes we want to see how a page has changed over time, or know when a page disappeared. Using the Wayback Machine can help you do that.

Here’s how that works. Go to the Wayback Machine and search for a page or site. Here we will search for the front page of the White House site:

![Wayback Machine](https://archive.org) whitehouse.gov

Explore more than 279 billion web pages saved over time

Figure 83

The Wayback Machine does not archive every page, but they do archive a lot of them. Whether a page is archived will often depend on whether a page was heavily linked to in the past, or if it was published by a site that the Wayback Machine tracks. In the case of the White House, of course, both these things are true and we have a near perfect history of the site.
Figure 84

Let’s go back in time all the way to 1999. When we select 1999, we see a calendar. Each circle indicates a snapshot made of the site. The green and blue indicate whether the page was a “redirect” – an issue beyond the scope of this article.

Click on a date to see a “snapshot” of the page on that date. Here we see a snapshot of the site from January 1999, at the tail end of the Clinton administration.
Sites will be browsable, to some extent, so go ahead and click on the links. Advanced functionality, such as search interfaces and interactive content, will usually not work.
FINDING OUT WHEN A PAGE WAS PUBLISHED USING GOOGLE

Many pages will tell you the date they were published. But some pages don’t give publication dates, and some can’t be trusted.

Take, for example, this story from fake site ABCNews.co (a hoax site that attempts to look like an ABC news site).

Figure 86
You’ll note that the publication date was November 11.

That’s what the site looks like today. But we can see what it looked like previously, courtesy of archive.org’s Wayback Machine.

Here’s what it looked like in March, sporting a publish date of March 24:
Figure 87

Here it is in June, sporting a date of June 16:
Figure 88

And in September, it sported a date of September 11:
Hoax sites often do this date incrementation to increase the share rate on older stories. People are more likely to share things if they believe they are breaking news and not yesterday’s story.

So how do we get some sense of when this story was first published?

We can’t get there exactly but we can often use Google to get close. Google stores the date of the first time it indexed a page–on popular sites this date is usually within a couple days of the true publish date (on unknown sites it is much less reliable).

To get Google to show the indexed date of a page, you’ll need to do two things:

- Set up a search that will only return that particular page by using the “site:” search term
- Trigger display date by setting a date range that ends with the current day.

Here’s what that looks like in this case:

As you can see, we’ve taken the URL of the page and entered the following as the search term:

site:abcnews.com.co/donald-trump-protester-speaks-out-i-was-paid-to-protest/

Then we’ve used date filtering to create a filter that doesn’t exclude anything (its date range is all possible dates), but triggers this sort of date display in Google.
Again, this is not a rock-solid publication date, but we can say that there was some content at this URL at this date, and in most cases, with a URL like this, that means the story was up by then.
Students often overestimate how many times the average paper gets cited. Students will look at a paper with 40 citations and say, “can it really be trusted with only 40 citations?” Most papers that get cited even a few times are legitimate papers (vs. junk). In all fields, 40 citations indicate a paper that has had a lot of expert eyes on it. That is the point of citations in source verification work. It is not necessarily about the quality of the paper. You need expertise to assess that, and a paper with 100 citations is not necessarily better than paper that has 10.

What citations show you, is that experts have read a certain work and found the work worthy of discussion. More citations do not mean more quality, but they do mean more expert eyes have probably looked at it and found it worth either agreeing or disagreeing with in public.

If you still want to know averages, here is a list of citation averages from 2011. Note that citations follow a power law, and any average here is far above the median.
36.

USING GOOGLE BOOKS TO TRACK DOWN QUOTES

Did Carl Sagan say this?

Charles Bergquist
@cbquist

Suspecting Carl Sagan had either a time machine or a crystal ball.

Ok, probably the time machine.

I have a foreboding of an America in my children’s or grandchildren’s time — when the United States is a service and information economy; when nearly all the manufacturing industries have slipped away to other countries; when awesome technological powers are in the hands of a very few, and no one representing the public interest can even grasp the issues; when the people have lost the ability to set their own agendas or knowledgeably question those in authority; when, clutching our crystals and nervously consulting our horoscopes, our critical faculties in decline, unable to distinguish between what feels good and what’s true, we slide, almost without noticing, back into superstition and darkness.

Quotes on the internet are some of the most commonly faked content. People misattribute quotes to give them significance, or fabricate tendentious quotes to create controversy. For some examples of fact-checking historical quotes, check out Quote Investigator.

If we know that Carl Sagan is an author of many books, rather than start in Google or DuckDuckGo’s general search, we might start in Google Books, which will likely get us to the source of the quote faster. Even if we cannot find the source, we might find someone quoting this in a book from a major publisher, which is likely to have a more developed fact-checking process than some guy on Twitter.

So we go to Google Books and we pick out a short snippet of unique phrasing. I’m going to choose “clutching our crystals and nervously consulting.”
Down there at the bottom, the fourth result, is a book by Carl Sagan. It is from 2011, but don’t be fooled by this date; this is just the date of the edition indexed here. Let’s click through to the book to check the quote and sort out the date later.

Clicking through the book, we find that the quote is accurate. More importantly, the surrounding context reveals that this quote is not being taken out of context. Sagan was truly worried about this issue. His prediction was that a sound bite obsessed media, combined with a sort of celebration of ignorance, would drag us backwards. He understood that the world was becoming more difficult while the communication of ideas was simultaneously becoming more shallow.
United States is a service and information economy; when nearly all the key manufacturing industries have slipped away to other countries; when awesome technological powers are in the hands of a very few, and no one representing the public interest can even grasp the issues; when the people have lost the ability to set their own agendas or knowledgeably question those in authority; when, clutching our crystals and nervously consulting our horoscopes, our critical faculties in decline, unable to distinguish between what feels good and what's true, we slide, almost without noticing, back into superstition and darkness.

The dumbing down of America is most evident in the slow decay of substantive content in the enormously influential media, the 30-second sound bites (now down to 10 seconds or less), lowest common denominator programming, credulous presentations on pseudoscience and superstition, but especially a kind of celebration of ignorance. As I write,

You can find out the original publication date of this work a number of ways. There’s a “more versions” option on the Google Books interface. You could go look for the book’s article on Wikipedia, as they will usually give you the publication date. But the easiest way is usually to turn to the front pages of the book and find the date, just as you would with a physical book.

A Ballantine Book
Published by The Random House Publishing Group

Copyright © 1996 by Carl Sagan

All rights reserved.


The copyright page from Carl Sagan's book.
Grassroots political efforts emerge from the bottom-up, with small local groups banding together to put pressure on city, county, state, or federal government to take (or oppose) specific action. They are “people-powered,” usually relying on volunteer labor and small donations from local people and organizations. In the age of social media, the phrase “grassroots” has also been applied to national movements that start by a small group of citizens organizing online.

Being “grassroots” is not a technique limited to Republicans or Democrats. The Tea Party revolts against President Obama’s health care plan, for example, had many grassroots elements, being organized on the local level by loosely connected people and local organizations. Moms Demand Action, a gun control advocacy group, was started when a stay-at-home mother was shocked by her son’s response to the Sandy Hook school shooting. She put up a Facebook page to organize action, and slowly built a movement.

Citizens tend to look more favorably upon people-powered, local politics rather than corporate-funded initiatives funded by people from somewhere else. The desire to portray corporate and non-local efforts as local has led to a practice called astroturfing, where large corporations or rich individuals use “front groups” that look like local groups of activists, but are funded and organized primarily by national corporations or rich individuals from elsewhere.

When deciding whether an organization is astroturfing, consider the following:

- Who funded it? (a corporation, national foundation, or local money)
- Who founded it? (if founded locally, by whom)
- What interest would that group have in the action or initiative proposed? (financial or related to larger social concerns)

There is a bit of a sliding scale for what qualifies as astroturfing. A locally founded initiative that receives primarily national money is a little less astro-turfy than an organization founded directly by a corporation. An initiative that receives money from a foundation dedicated to a larger social goal (such as elimination of poverty) is less astro-turfy than a corporation spending money to boost its stock price or get rid of regulations that constrain it. In general, what is most important is whether the organization’s reality matches the story that they are publicly telling.
SEITRCHING TV TRANSCRIPTS WITH THE INTERNET ARCHIVE

The Internet Archive, is building a digital library of Internet sites and other cultural artifacts in digital form. In 2009, the Internet Archive made selected U.S. television news broadcasts searchable by captions in their TV News Archive. This service allows researchers and the public to use television as a citable and shareable reference.

For example, a search for Canada’s Prime Minister Trudeau and US President Donald Trump contains clips from various news organizations showing the two leaders together or one of the leaders mentioning the other at a news conference. The site provides many ways to filter content – program, creator, year, topic, etc. There are also several ways to sort the results. Another good tool to have in your fact-checking toolbox.
Occasionally when you search for an answer to a question on Google, you will not only find websites, but you may also find a “knowledge panel” that appears to have what search expert Danny Sullivan calls the “One True Answer,” an answer that appears on a knowledge panel on top of the results.

Sometimes Google pulls an answer from a source algorithmically. For example, in response to “How many men landed on the moon?,” this panel answers “12 men,” citing a Quora article.

Sometimes Google does not pull out the answer but makes the answer apparent in the blurb or headline of the card, as in this answer to the query, “last person to walk on the moon”.

![Google search result for how many men landed on the moon](https://www.google.com/search?q=how+many+men+landed+on+the+moon&ie=utf-8&oe=utf-8&client=ms-2&rls=org.mozilla:en-US:official&source=lnms&tbm=isch&sa=X&ved=0ahUKEwi436W53fjXAhUGpTcKHafCDh4Q_AUICCgB#imgrc=QJc5KwW-HGy5-n)
This function of Google can be useful, but it malfunctions frequently enough that it should not be trusted without verifying the source and context of the answer. There are two major problems: false simplicity and false (or non-standard) information.

**False Simplicity**

Consider this question – how many apostles are there in the Christian tradition? Google tells you, via a panel, even pulling out the number, thereby making it look decidedly objective – there are twelve!
If you click through to that Quora question, you will find that it answers a much more specific and simpler question – how many original apostles did Christ have (according to tradition). And for that answer they are correct; including Judas, there are twelve.

But according to tradition, when Judas dies Matthias becomes an apostle, so that’s thirteen. Then, Paul is an apostle, so fourteen. And Barnabas, Timothy, and James. The truth is that this answer is pretty debatable. It is certainly not twelve, and some versions of the Bible refer to 25 different people as “apostles.”

These numbers, which are varied, come from various Christian traditions. Many historians, on the other hand, see the twelve apostles as a creation of the early Church, that had no reality or significance during the lifetime of the historical Jesus and was later “retrojected” into the Gospels.

The whole question of how many apostles there were and who they were is inextricably bound up with complex questions of religion, history, and 1st century power struggles about who counted in the early church and who didn’t.

Anyone who investigates religious or historical sources for an answer to this question will be confronted by a variety of opinions. Google’s panels, however, are oblivious to this kind of complexity and present a simple numerical answer where no simple answer actually exists.

**Misleading Highlights**

Google uses programming to try and highlight relevant answers in the blurb (below), but the highlighting is confusing. When asked how old Lee Harvey Oswald was when he shot Kennedy, Google highlights 18, 24, and 22.


In reality, the answer is 24 years old, though a quick glance at this might have you thinking 18 or 22.
Blatant Misinformation

Sometimes the panel presents blatant misinformation. Often this material is the product of highly politicized areas or of conspiracy-believing communities, which tend to rank highly in Google search results.

Take for instance this search, where we ask Google which presidents were in the Ku Klux Klan. The Google panel provides what seems to be a definitive answer. There were five!

As Case University Western history professor Peter Schulman points out, this isn’t even remotely true. None of these presidents were members of the Ku Klux Klan (as far as we know), and if you click through to the article, you’ll find the source is a Nigerian newspaper of uncertain stature that references a book by David Barton, a nationalist known for self-publishing dubious works of historical revisionism.

There are numerous examples of similar behavior. Adrianne Jefferies at The Outline details some more bad snippets, including this one claiming Obama is planning for martial law (complete fiction).
Google will also tell you that Lee Harvey Oswald didn’t assassinate John F. Kennedy, despite the overwhelming evidence to the contrary:

Google will also tell you that Lee Harvey Oswald didn’t assassinate John F. Kennedy, despite the overwhelming evidence to the contrary:

Confirmation Bias and Bad Snippets

A lot of times Google is just bad. But sometimes bad answers are often the result of asking questions in ways that tap into the language or concerns of pseudoscience, conspiracy theory, or fringe beliefs. For example, some people have problems with monosodium glutamate, a food additive that triggers an allergic reaction in a small portion of the population. If you search on a phrase likely to be found in the medical literature like “msg sensitivity,” you get a fairly reliable result. Healthline is a recognized provider of reliable health information.
This changes if you use the language of fringe groups that believe the medical community is suppressing a link between MSG and a variety of neurological disorders. Note the results you get when you type in ‘msg dangers’:

The blurb contains words such as brain damage, Alzheimer’s disease, and learning disabilities. If you look up the site, mercola.com, you will find that it is run by a physician who has been repeatedly warned by the FDA to stop making false claims.
Our Advice

In general, simply treat the Google panel (“one true answer”) as you would any other top search result. Despite Google’s claims to the contrary, it is not significantly more or less reliable than an average source. Click through, trace the claims on the page to a source, and investigate the source. Never trust the result without validating the source of the claim.
If you are looking to hone your fact-checking skills, you may want to find highly viral stories. Your own Facebook and Twitter feeds are good sources for such stories.

Sometimes, you will want to get outside your filter bubble and see the stories that other folks are sharing. There are a number of tools you can use to find viral stories. For example, Buzzsumo is an option that is simple to use. Go to Buzzsumo.com.

Put in a search term, like “cancer.” Buzzsumo will return the most shared stories on the topic of cancer. Filter them by date, and we can look at the stories for the past week.
Facebook engagements is not purely about shares and encompasses other actions as well. It is a good metric of how viral the story is.

The free version of Buzzsumo only lets you view the top results and limits the number of searches you can perform per day; however, it is often enough access to enable you to find an interesting story to fact-check. The “Cancer Cure Genius Silenced by Medical Mafia” result is interesting. Its inflammatory language is a good indicator that the claims in it are likely to be overstated.
Free version of Buzzsumo search results
FINDING OUT WHO OWNS A DOMAIN

You may want to know who is behind a domain. In the past, this used to be relatively easy to find out. When a person bought a domain, their name was put into a “registry,” which is similar to the “phonebook of domains.” To find out who owned a domain, you would just go and look it up, using a service we discussed in an earlier section called *WHOIS*.

Unfortunately, it is now more complicated. People who had their email addresses and names in the “domain phonebook” would get spam email, or the information displayed on the registry would be used to try to hack their site. And many people, such as political dissidents, had good reason for not revealing their names. As a consequence, a lot of the “registrars” started offering masking services, which hide the owner of the domain.

Now if you want to find out who owns a domain, *WHOIS*-type services are a good first stop, even though they will usually fail for smaller sites.

To look up domain ownership, we recommend a tool called *Domain Dossier*. Go to the site and type in your root URL and check all the checkboxes.

When the identity is not masked, you will be able to see the owner of the domain. The first place to look is “Registrant Name” and “Registrant Organization”:
Occasionally, you may not get a useful name from the record, but the address might be telling. If the name is masked in *Domain Dossier*, you’ll get a record that looks like this:

**Domain Whois record**

Queried `whois.internic.net` with "`dom protrump45.com`"...

```
Domain Name: PROTRUMP45.COM
Registry Domain ID: 2083681824_DOMAIN_COM-VRSN
Registrar WHOIS Server: whois.NetworkSolutions.com
Updated Date: 2017-01-28T00:10:12Z
Creation Date: 2017-01-28T00:10:11Z
Registry Expiry Date: 2018-01-28T00:10:11Z
Registrar: Network Solutions, LLC.
Registrar IANA ID: 2
Registrar Abuse Contact Email: abuse@web.com
Registrar Abuse Contact Phone: +1.8003337680
Domain Status: clientTransferProhibited https://icann.org/epp#clientTransferProhibited
Name Server: NS0.WIXINS.NET
Name Server: NS1.WIXINS.NET
DNSSEC: unsigned

URL of the ICANN Whois Inaccuracy Complaint Form: https://www.icann.org/wicf/

>>> Last update of whois database: 2017-06-08T06:00:04Z <<<
```

Queried `whois.networksolutions.com` with "`protrump45.com`"...

**No match for "PROTRUMP45.COM".**

You may also see the name of a masking service, such as “Domains by Proxy”:

```
Registry Registrant ID: Not Available From Registry
Registrant Name: Registration Private
Registrant Organization: Domains By Proxy, LLC
Registrant Street: DomainsByProxy.com
Registrant Street: 14455 N. Hayden Road
Registrant City: Scottsdale
Registrant State/Province: Arizona
Registrant Postal Code: 85260
Registrant Country: US
Registrant Phone: +1.4806242599
Registrant Phone Ext: 
Registrant Fax: +1.4806242598
Registrant Fax Ext:  
```
In this case, the registrant is not from Arizona and not named “Domains by Proxy”. That is just the masking service. It is important to note that masking is common enough these days and should not cause suspicion.

While domain owners can hide their names, they cannot hide the date the domain was registered. As we will discuss in another chapter, this is often useful information. By looking at the domain registration date, you can often get a sense of whether a site has a long history behind it or if it has been spun up for a specific purpose.
AVOIDING CONFIRMATION BIAS IN SEARCHES

Was 9/11 a hoax? Let’s find out. We type in ‘was 9/11 a hoax’ and we get:

![Google search results for was 9/11 a hoax]

Well, look at that. Not only the top result says that the attack on 9/11 was faked–the top five results do. To the untrained eye it looks like the press has been hiding something from you.

But of course the 9/11 attacks were not faked. So why does Google return these results?

The main reason here is the term. The term “hoax” is applied to the 9/11 attacks primarily on conspiracy sites. So when Google looks for clusters on that term (and links to documents containing that term), it finds that conspiracy sites rank highly.

Think about it: reputable physics journals, policy magazines, and national newspapers are not likely to run headlines asking if the attacks were a hoax. But conspiracy sites are.
The same holds true even for more benign searches. The question, “Are we eating too much protein” has Google return a panel from the Huffington Post (now HuffPost) and a website from a vegan advocacy group.

To avoid confirmation bias in searches:

- Avoid asking questions that imply a certain answer. If I ask “Did the Holocaust happen?,” for example, I am implying that it is likely that the Holocaust was faked. If you want information on the Holocaust, sometimes it’s better just to start with a simple noun search, e.g. “Holocaust,” and read summaries that show how we know what happened.
- Avoid using terms that imply a certain answer. As an example, if you query “Women 72 cents on the dollar” you’ll likely get articles that tell you women make 72 cents on the dollar. But if you search for
“Women 80 cents on the dollar” you’ll get articles that say women make 80 cents on the dollar. Searching for general articles on the “wage gap” might be a better choice.

- Avoid culturally loaded terms. As an example, the term “black-on-white crime” is a term used by white supremacist groups, but is not a term generally used by sociologists. As such, if you put that term into the Google search bar, you are going to get some sites that will carry the perspective of white supremacist sites, and be poor sources of serious sociological analysis.

- Plan to reformulate. Think carefully about what constitutes an authoritative source before you search. Once you search you’ll find you have an irrepressible urge to click into the top results. If you can, think of what sorts of sources and information you would like to see in the results before you search. If you don’t see those in the results, fight the impulse to click on forward, and reformulate your search.

- Scan results for better terms. Maybe your first question about whether the Holocaust happened turned up a poor result set in general but did pop up a Wikipedia article on Holocaust denialism. Use that term to make a better search for what you actually want to know.
Promoted tweets are real tweets, but they do not reach you because they were shared by the people you follow. They reach you because the author of the tweet paid *Twitter* money to put it in your feed.

Here’s an example of a promoted tweet, asking you to “Tweet your Senators” about the dangers of drug importation:

Promoted tweets are not necessarily untrue, but they should be treated the way one would treat a commercial. In this case, we look to see what organization has posted the tweet.
That leads us to their webpage and organization name: The Partnership for Safe Medicines.
And a little bit of investigation takes us to a page on the NPR site that shows this organization has ties to Big Pharma:
While none of these means the claims of the organization are wrong or false, it is a worthwhile perspective to have before you decide to retweet the tweet or not. Treat promoted tweets with suspicion. Someone is paying money to influence you, and it’s best to know who before retweeting.
FINDING OLD NEWSPAPER ARTICLES

While more recent news articles are available from both Google’s and Bing’s news search tabs, older news can be more difficult to retrieve. Many options for retrieving old news entail paying a subscription fee or per article cost, which is a bit expensive for a person just checking up on a story. In this section, we will show you how to use news archives to check on the existence of articles at no cost.

A Sample Problem

President Trump claimed the investigation to see if his campaign had colluded with Russia was a “witch hunt.” Shortly afterward, this snapshot of an article appeared in a news feed:

Figure 119
By now you should know it is easy to fake something that looks like a snapshot of an old headline. So how do we find out if this article actually ran?

Our first instinct might be to go to the *Washington Post* to see if they have this article. That’s not a bad instinct, but in this case the headline clearly ran somewhere else other than the *Post*. The *Washington Post* does not tag its own articles as coming from the “*Washington Post*.” This particular headline ran in another paper.

So we want to do a broad search across many historical American papers. When reporters do this, they most often use tools such as *LexisNexis* and *ProQuest*, which are usually unavailable to average people.

We will have to make do with sources that are searchable from the web. There are three major web searchable archives:

- *Google’s Historical Newspapers*: news.google.com/newspapers
- *Newspapers.com*: newspapers.com
- *Newsbank’s Newspaper Archive*: newspaperarchive.com

*Google* offers complete articles. The other two offer snippets unless you pay them money, but snippets are enough for this sort of task.

So we construct our search. It is just a variation on the “site:” syntax we’ve used elsewhere.

Nixon Sees Witch Hunt (site:newspapers.com OR site:news.google.com/newspapers OR site:newspaperarchive.com)

And we get back a time-stamped result from the *LA Times*, with a date (in 1973) that looks promising:
Figure 120

Note the column on the left says, “Nixon sees witch-hunt Sears insiders say.” Why does the word, “Sears” appear there?

It becomes evident when we click through and look at the page:
Figure 121

You can see above we have circled the headline. The free version only offers this blurry “thumbnail” image of the page, but it’s enough to spot the headline. It also makes obvious where the “Sears” came from. The text here was automatically generated by computer and must have included the Sears ad next to it as part of the headline.

If we scroll down the page, we can see enough to confirm that this article was correct, even though the automatic character recognition has messed up a lot of the words:

Part l-A-Sun., July 22, 1973 I Nixon Sees ‘Witch-Hunt; Sears Insiders Say Prices Effective through Tuesday, July 24 BY BOB WOODWARD and CARL BERNSTEIN Thft Washington Post WASHINGTON President Nixon and his top aides believe that the Senate ‘Watergate hearings are unfair and constitute a “political witch-hunt,” according to White House sources. The sources, said, that the President .in recent weeks had expressed bitterness and deep hostility toward the two-.morith-old proceedings.

We have enough here to say that this ran in the LA Times in July 1973. And if we really wanted to see a clean
version of the article, we could subscribe to the service and grab a better image, which may be what the original tweeter did.

Checking Cited Headlines

Here’s another paragraph, this time from the \textit{New York Times}, that claims the \textit{LA Times} ran a derogatory headline when the first female commercial pilot at a major airline got her wings.

There were no female pilots at the biggest airlines until 1973, when American Airlines hired the first, Bonnie Tiburzi Caputo. In a reminder of how times have changed, that news was reported in The Los Angeles Times under the headline, “Airline Pilot to Fly by Seat of Panties.”

The \textit{New York Times} is a very reliable paper, and in this case we probably do not need to check the article title. But let’s try anyway with the same sort of search as above:

Airline Pilot to Fly by Seat of Panties (site:newspapers.com OR site:news.google.com/newspapers OR site:newspaperarchive.com)

Note that because the optical character recognition sometimes transcribes things incorrectly, we don’t put quotes around the search phrase, at least at first. When we put it in, we are in luck; we can see the headline in the blurb:

![Image of search results]

Figure 122

We might also search for a type of headline. For instance, a key point of the critics of global warming is the statement that scientists in the 1970s feared “global cooling” instead of global warming; the point being that...
the global warming scare is one in a long series of bad guesses to be later thrown away. Can we compare the number of global cooling and global warming stories in the 1970s?

We execute a search for:

```
global cooling (site:newspapers.com) 1975
```

and we get an article from 1975, which talks of some sensationalist claims of a coming ice age. But when the reporter talks to a climatologist, the tone is different:

But Lawson prefers to speak in terms of the following probabilities: —In the long run, over thousands of years, there is probability of an ice age. —In the next few decades, there is a probability of a warming trend. —In the next few years, the probability is that global cooling will continue downward to 19th century levels.

(Note: For some reason newspaper archive searches react badly to date filters, which is why we just put 1975 in plain text.)

If we search for “global warming” in 1975, we get this quote in the January 29, 1975 edition of the Orlando Sentinel from a government scientist:

“After the next decade or so will come a warming trend, both because of increased CO2 in the atmosphere and thermal pollution by power plants and so on. In the 21st century, man’s activities will predominate over nature.” J. Murray Mitchell, senior research climatologist, Environmental Data Service, National Oceanic and Atmospheric Administration.

While one would need much more evidence to settle the question of whether scientists on the whole feared global cooling or global warming in the 1970s, it is clear enough that many scientists expected warming due to man’s activities even then. If you’re looking at sharing an article that says that “cooling” was the big 1970s worry, you might want to get more information about it before reposting.