Prologue

Let's play a little game. Try to guess if the following statements are true or false (no peeking below for the answers!). The virus that causes COVID-19 came from outer space. King Charles III of Britain is Count Dracula's cousin. John Kellogg invented cornflakes to prevent people from having impure sexual thoughts. The US military hides information about UFOs from the public. The average person accidentally swallows about eight spiders per year in their sleep.

Are you ready? Let's see how well you did.

COVID-19 most likely didn't come from outer space. This may seem obvious, but in 2020 an apparently dead serious group of scientists claimed that the virus may have "arrived via a meteorite […] that struck Northeast China on October 11, 2019" (E. J. Steele et al., 2020). Who knows if their claim will be proven true at some point, so let's put this one in the "maybe" pile. King Charles does indeed appear to be a distant relative of Vlad III of Wallachia, better known as Vlad the Impaler, who inspired Bram Stoker's Dracula (Beever, 2022; CBS News, 2011). The claim that John Kellogg invented cornflakes as an "anti-masturbatory morning meal" is widespread (Mayyasi, 2016; Soniak, 2023), but the fact-checking website Snopes rates it as "mostly false": Kellogg designed cornflakes to be easy to digest, pre-prepared, and healthy, although he did recommend a bland diet to discourage masturbation (it's just that there's not much evidence, apparently, that he invented cornflakes for this reason). It's a fact that the US military hides information about UFOs from the public (Hopkins & Snyder, 2022), but this statement is incomplete: UFOs (Unidentified Flying Objects or Unidentified Anomalous Phenomena, UAPs, as the military calls them nowadays) are commonly observed, but officials are wary of disclosing too much information about them for national security reasons (e.g., because they don't want the public know about secret military technology); the *implied* claim that

was not made explicit in this statement, however, is unlikely to be true as of this writing.[1]

And finally, people don't accidentally swallow many spiders in their sleep, although this is a widespread myth.[2] This is actually a brilliant example of how easy it can be to spread misinformation, but not in the way you might think. In 2001 David Mikkelson of *Snopes* published an article debunking the spider myth (Mikkelson, 2001). He explained how unlikely it is for spiders to crawl into people's mouths by accident, and even found the myth's original source: a 1993 article in the magazine *PC Professional* by Lisa Birgit Holst about the dangers of internet misinformation. In turn, Holst had taken the claim from a 1954 book about insect folklore that listed a collection of common misbeliefs. Ironically, Mikkelson explained, Holst's article became so popular that the spider myth became one of the most widely circulated pieces of misinformation on the Internet. The Snopes article went viral, and other websites followed suit by publishing their own debunking articles. Not only the claim itself but also its correction became widespread. Many of these websites uncritically copied Mikkelson's explanation of the myth's origin, but there was a catch: it was completely made up. A Swedish YouTuber named LEMMiNO (2016) conducted an exhaustive investigation in search of Holst's article in *PC Professional*, but came up with nothing. He then realized that he'd been tricked: Lisa Birgit Holst is an anagram for "this is a big troll." Mikkelson had made her up to prove the point that many people uncritically accept information without fact-checking it (Snopes, 2023). The trolling worked brilliantly: in their eagerness to dispel what they thought was a commonly believed piece of misinformation, debunkers simply copied Mikkelson's explanation without doing any further research.

If you didn't know all of this, you're not alone: it can be very difficult to tell fact from fiction (Channel 4, 2017). Some lies are so well crafted that they're almost impossible to debunk, and some of the most cleverly designed misinformation contains a kernel of truth. Take the UFO claim from earlier: it's objectively correct, but phrased in such a way that it's easy to infer a wrong conclusion. Objective truths and the subjective interpretation of them can sometimes be hard to disentangle.

The above claims are all relatively benign, but misinformation isn't always harmless. The Protocols of the Elders of Zion, a conspiracy theory invented by the Russian secret service under Tsar Nicholas II, is still influential today as a justification for anti-Semitic beliefs (Bronner, 2007; Whitfield, 2020). The spread of vaccination myths has sometimes been associated with a drop in

[1] **ALIENS!** Although some military officials have said publicly that "the U.S. has retrieved craft of non-human origin" (Gabbatt, 2023; Kean & Blumenthal, 2023), with one official testifying in front of Congress that he was informed of "a multi-decade UAP crash retrieval and reverse-engineering program" (Becket, 2023). So maybe we're wrong. Who knows.

[2] We don't have numbers on how many spiders are swallowed on purpose.

vaccine uptake (Larson, 2020; Motta & Stecula, 2021). And false and misleading news stories about nuclear energy can influence people's risk perceptions around nuclear technology, which might hamper informed debate about how to combat climate change (Ho et al., 2022; Thakur & Ward, 2019). But what exactly *is* misinformation? To what extent does misinformation take on problematic proportions for society? Who believes and shares misinformation, and why? How does the rise of modern technology such as social media impact how misinformation is created and spread? And what solutions do we have available, how well do they work, and do these solutions themselves not pose potential risks? These are some of the questions that we tackle in this book.

Research on the spread of, belief in, and consequences of misinformation has long been popular (H. M. Johnson & Seifert, 1994; Lewandowsky et al., 2005; McGuire, 1961; Price & Hsu, 1992), but really took off around the 2016 US presidential elections. If you look up "misinformation" on Web of Science, you'll find that the number of publications on the topic rose from 43 per year in 2000 to 231 in 2015, and to a whopping 2,249 in 2022. While it's impossible to tackle the entirety of this sprawling research field on these pages, we've done our best to cover the most important scientific insights into various aspects of the misinformation problem. To do this, we've tried to incorporate as many viewpoints as possible. We of course have our own thoughts and ideas, and we discuss these where relevant, but we've tried to be as nuanced as we could when discussing the various perspectives and research approaches. For instance, Chapters 3 (which asks whether misinformation is a problem), 4 (which examines why people believe and share misinformation), and 5 (which tackles echo chambers and filter bubbles) explicitly discuss competing perspectives in detail. We hope we've done each of these perspectives justice.

We have divided this book into three parts. Part I (Chapters 1–3) sets the stage and covers the necessary background information. Chapter 1 delves into the complexities of defining misinformation: as the above examples showed, it's often difficult to know if a piece of information is objectively true or false. Instead, misinformation is oftentimes contextual: truth value might be absent altogether, a statement might leave out relevant information, or the framing might be such that a true claim is nonetheless misleading. We explain these complexities using illustrative examples, categorize the different definitions that researchers have used over the years, and present our own working definition of misinformation which we will use throughout this book. Chapter 2 is about the history of misinformation, starting with its evolutionary origins and deception in the animal kingdom. We then discuss how information was produced and spread throughout human history, focusing especially on the role of technological innovations. We begin this story in the time before Gutenberg's version of the printing press came about around 1440 AD, before moving on era of the printing press (until 1920 or so), the mass media era (1920–1990s), and finally the internet age (from when Al Gore invented the

Internet until the present day). Chapter 3 tackles whether the spread of misinformation poses a problem for society. The answer to this question might sound intuitive (especially coming from two people who are writing a book about it), but it's not: there are many nuances here that are worth exploring in detail, and some scientists have argued that the misinformation problem is at least somewhat overblown.

Part II (Chapters 4–5) covers the belief in and spread of misinformation. Chapter 4 addresses these questions from a psychological perspective. It discusses the various theories that researchers have proposed and tested over the years about why people believe things that aren't true, and why they might share false or misleading information with others (e.g., on social media). We discuss the belief in and sharing of misinformation separately. This is because, although it might sound logical that people usually share misinformation *because* they believe it, this isn't always the case. Rather, misinformation belief and sharing are explained by somewhat (although not entirely) different mechanisms. Chapter 5 is about digital technologies, and whether the arrival of the Internet has ushered in an era in which it's uniquely easy for misinformation to proliferate. This chapter focuses on echo chambers and filter bubbles, and asks if social networks promote isolation from inconvenient ideas and viewpoints (more so than was the case before the Internet).

Finally, Part III (Chapters 6–8) is concerned with how to counter misinformation. Chapter 6 looks at how misinformation is regulated by governments and other legislative entities. Misinformation has become of increasing interest to lawmakers around the world (both in democratic and not so democratic countries). How have countries such as the US and UK, and supranational bodies such as the European Union, tried to regulate how misinformation (and related content such as hate speech) is consumed and spread on social media platforms? How comprehensive is this new legislation? And, importantly, what are the risks of creating new laws (on top of already existing limits to speech such as libel, slander, and fraud), for example when it comes to inspiring autocratic countries to crack down on political opposition under the guise of fighting "fake news"? Chapter 7 explores anti-misinformation measures that are targeted at individuals, focusing on interventions that either tackle susceptibility to misinformation or seek to reduce the sharing of it. We cover four different types of interventions: *boosts* (which aim to build new skills and competences or foster existing ones, such as media literacy and critical thinking programs, as well as "prebunking"); *nudges* (subtle, nonintrusive interventions that incentivize positive behavior, for example reducing the sharing of misinformation on social media); *debunking* (correcting misinformation after it has already spread, for example through fact-checking); and the *(automated) labelling* of various kinds of problematic content. We discuss the evidence behind of these interventions from lab and field studies, as well as what we don't know (yet). And finally, Chapter 8 is about our own research

program, much of which has focused on "fake news" games and videos that are grounded in a framework from social psychology called "inoculation theory." The chapter details how this research program came about, as well as some of the insights we believe we've arrived at over the years. We also wanted to emphasize some of the shortcomings and nuances of this research, and discuss the (in our view) important implications of these nuances for researchers and policymakers seeking to understand and counter misinformation.

Each chapter is structured in a way that we're told is common in publishing, with an introduction, a main body, and a conclusion. Where relevant, we've added footnotes at the bottom of the page with additional information. To avoid burdening the reader with excessive footnotes, we've opted for what we believe is a unique approach in scientific writing: all information that is of critical importance for an accurate understanding of each chapter can be found in the footnotes, whereas any contextual and other nonessential information is provided in the main body. You can therefore safely read only the footnotes and nothing else.[3] However, we recommend reading the whole book including the main text, as we've worked pretty hard on it and would be disappointed if you skimmed it.[4]

Attentive readers might ask: "Hey Sander, didn't you just write a book about misinformation? Why do you need a new one?" To answer these questions: yes, Sander did publish a book called *FOOLPROOF: Why Misinformation Infects Our Minds and How to Build Immunity* (2023). As Jon was tasked with writing the initial draft of the chapters in the book you have in front of you, he has heroically avoided reading *FOOLPROOF* to prevent himself from being influenced by it. So although there is overlap between the two books, there are important differences in scope, topics of discussion, and especially the conclusions.[5] Most importantly, we've done our best to incorporate as many different perspectives as possible. Only in the last chapter (Chapter 8) do we address our own work in detail, but in a (what we hope to be) appropriately self-critical manner: over the years, some people have disagreed with our approach to misinformation intervention design. Much of this disagreement has been reasonable and well intended, although not all of it (for the last time: we're not lizard people[6]). Some of these critiques have prompted us to make substantial changes to how we conduct our research and design our interventions, which we hope will be informative for people looking to learn more about how misinformation works, and what to do about it.

[3] Just kidding, the main text is important too. The large print giveth and the small print taketh away (Waits, 1973).

[4] As a former colleague of ours (Elif Naz Çoker) likes to say: we're not *mad* scientists, just disappointed ones.

[5] An even more attentive reader might ask: "how do you know this, Jon, if you haven't read *FOOLPROOF*?" Point taken.

[6] Although that is what a lizard person would say, isn't it?

Finally, we wrote this book with an audience in mind that is interested in misinformation, but doesn't necessarily have a psychology, behavioral science, or political science degree (or maybe not yet). We've avoided scientific jargon as much as possible, and explain numerical information (e.g., in graphs) in nontechnical terms where relevant. However, we haven't compromised too much on complexity: some topics (such as how echo chambers work, see Chapter 5) are not always intuitive to grasp, but they're critical for a comprehensive understanding of misinformation. We've intended for this book to serve both as a textbook for students taking courses on mis- or disinformation at the undergraduate and postgraduate level, and as a resource for people outside of academia (e.g., policymakers, educators, or people who work for fact-checking organizations, but also interested nonexperts). We've done our best to incorporate the most recent and most robust scientific research in the field, and to represent it fairly and reasonably. Most of all, we hope you'll enjoy reading this book.

PART I

SETTING THE STAGE

1

Defining Misinformation

1.1 INTRODUCTION

Most people who regularly use the Internet will be familiar with words like "misinformation," "fake news," "disinformation," and maybe even "malinformation." It can appear as though these terms are used interchangeably, and they often are. However, they don't always refer to the same types of content, and just because a news story or social media post is false doesn't mean it's always problematic. To add to the confusion, not all misinformation researchers agree on the definition of the problem or employ a unified terminology. In this chapter, we discuss the terminology of misinformation, guided by illustrative examples of problematic news content. We also look at what misinformation *isn't*: What makes a piece of information "real" or "true"? Finally, we'll look at how researchers have defined misinformation and how these definitions can be categorized, before presenting our own definition. Rather than reinventing the wheel, we've relied on the excellent definitional work by other scholars. Our working definition is therefore hardly unique; we and many others have used it as a starting point for study designs and interventions. We do note that our views are not universally shared within the misinformation research community. We can therefore only recommend checking out other people's viewpoints with respect to how to define the problem of misinformation or related terms such as fake news, disinformation, and malinformation (Altay, Berriche, et al., 2022; Freelon & Wells, 2020; Kapantai et al., 2021; Krause et al., 2020; Lazer et al., 2018; Pennycook & Rand, 2021; Tandoc et al., 2018; Tay et al., 2021; Vraga & Bode, 2020; Wardle & Derakhshan, 2017).

1.2 FAKE NEWS, MISINFORMATION, DISINFORMATION, MALINFORMATION...

On the surface, "misinformation" seems easy to define. For instance, you might say that misinformation is "information that is false" or tautologically define

FIGURE 1.1 Example of a false but relatively harmless news headline (*The Onion*, 2012).

fake news as "news that is fake." Or, to use David Lazer's more comprehensive phrasing, fake news is "fabricated information that mimics news media content in form, but not in organisational process or intent" (Lazer et al., 2018, p. 1094). An example of a news story that would meet Lazer's definition is a hypothetical headline from www.fake.news that reads "President drop-kicks puppy into active volcano." This headline is false (that we know at least), and www.fake.news mimics news content but is not universally considered to be a trustworthy source of information (and presumably doesn't follow the ethical guidelines and editorial practices used in most newsrooms). Many would also agree that this headline may have been written with malicious intent in mind (assuming the authors weren't joking): If someone were to believe it, they would be left with an inaccurate perception of another person, in this case the president, which might inform their decision-making (e.g., leading them to vote for somebody else or disengage from the political process altogether). However, things aren't always this straightforward: Not only is it sometimes difficult to discern what is true or false, but true information can also be used in a malicious way, and false information can be benign and sometimes hilarious. Take, for example, the headline from Figure 1.1.

The Onion is an American satirical news site that publishes humorous but false stories which mimic regular news content. The rather wonderful part about this specific story is that the online version of the *People's Daily*, the Chinese Communist Party's official newspaper, apparently believed that it was real, and reposted the article along with a fifty-five-page photo reel of the

North Korean leader (BBC News, 2012). But despite the story being entirely false, it was benign, even though some people believed it (though we recognize some feelings may have been hurt when people found out that Kim Jong-Un *wasn't* named 2012's sexiest man alive). If anything, the ability to make fun of powerful people through satire is often seen as a sign of a healthy democracy (Holbert, 2013), and as far as satire goes the *Onion* story was rather mild.

At the same time, false information is not always benign, nor does it always try to mimic regular news content. Figure 1.2 shows an example of a Facebook post which got quite a bit of traction in 2014 and again in 2016.

The post is associated with #EndFathersDay, a fake hashtag movement started by members of the 4chan message board some time in 2014. Some 4chan users wanting to discredit feminist activists came up with a talking point that they thought would generate significant outrage and tried to get it trending on Twitter (Broderick, 2014). These kinds of artificial smear campaigns imitate the language and imagery of a group in order to bait real activists and harm their credibility. The image from Figure 1.2 was manipulated to make it look like there were women on the streets protesting for the abolition of Father's Day. But if you look closely, you'll see that the image was photoshopped. The demonstration where the original picture was taken was about something entirely unrelated. Nonetheless, this post is an example of how easy it is to manufacture outrage online using very simple manipulation tactics. Its creators didn't even have to bother setting up a "news" website to spread their content or mimic media content in form: All they had to do was photoshop a picture of a demonstration and spread it on social media.

To add another layer of complication, the *intent* behind the production of (mis)information also matters a great deal. It can happen that someone creates or spreads misinformation unintentionally (analogously to a virus spreading among asymptomatic people). For example, a journalist could write a news article fully believing it to be true at the time, only for the information in the article to later turn out false. Simple errors can and do happen to the best of us. Similarly, someone may share something on social media that they either erroneously believe to be true or don't believe but share anyway because they're distracted (Pennycook & Rand, 2021). Oftentimes, however, misinformation is produced intentionally. In February 2022, just after the start of the Russian invasion in Ukraine, Melody Schreiber at the *Guardian* noticed a drop-off in the activity of Twitter bots that were spreading misinformation about COVID-19 vaccines (Schreiber, 2022). The reasons behind this reduced activity were varied, but it appears that a significant number of Twitter bots that were spreading COVID-19 misinformation were run from within Russia. These bots went dormant for a while, but soon became active again to pivot their attention away from COVID-19 and toward the war in Ukraine. Building a Twitter bot and programming it to spread misinformation takes some time, effort, and money, and it's reasonable to assume that the people who created the bots intended for them to spread false and misleading