



Uio • University of Oslo

The Internet as Source of Epistemically Relevant Information

*Epistemic individualism versus epistemic
structuralism*

Christer Hølestøl

Supervisor: Associate Professor Sebastian Watzl

Master's Thesis in Philosophy

60 points

Department of Philosophy, Classics, History of Art and Ideas

Faculty of Humanities

University of Oslo

2022, Autumn Semester

Preface

First and foremost, I want to thank Sebastian Watzl, who was my supervisor from January 2021 to February 2023. His expertise, kindness, continued encouragement, and contagious enthusiasm have been a tremendous resource at all stages of the writing process. I also want to thank Kai Olav Ellefsen, Roger Antonsen, and especially Odd Erik Gundersen for pointing me to good resources about computer science. Gundersen liberally shared his time, expertise, and advice about writing about artificial intelligence. I also want thank Trine Syvertsen and Gunn Enli for discussions about epistemic responsibility. Finally, I want to thank my family, friends, and especially Silje Ratcharanarin Khuanroodee for their continued support and encouragement. The errors, of course, are all mine.

Abstract

Many of us go to the Internet to gather information relevant to our questions. Yet, much of the information on the Internet is irrelevant for answering them. How, then, can the Internet be a source of relevant information? How can it become a *better* source of relevant information?

My thesis explores two questions: What explains successful and unsuccessful information gathering online? And how can information gathering online best be improved? I distinguish between two positions: epistemic individualism and epistemic structuralism. Epistemic individualist accounts explain (un)successful online information gathering primarily in terms of Internet users, such as their open-mindedness (or lack thereof). I call this “epistemic individualism about explanation.” Relatedly, epistemic individualist accounts can also maintain that improvements to information gathering online should focus on Internet users—e.g., making them more open-minded. I call this “epistemic individualism about amelioration.” However, structural factors beyond users arguably impact information gathering online—e.g., how biased a tool is towards providing irrelevant information. I call accounts primarily emphasizing structural factors in explanations and improvements “epistemic structuralism about explanation” and “epistemic structuralism about amelioration,” respectively. The thesis assesses whether epistemic individualism or epistemic structuralism provide the best explanations and suggestions for improving information gathering online.

In Chapter 1, I take an account by Richard Heersmink as a starting point. Interpreting it as an epistemic individualist account, I take the main claim to be that a user’s good character traits are crucial for successfully using search engines to gather information online. However, drawing upon work by C. Thi Nguyen, I argue echo chambers can make members with good character traits engage in bad information gathering practices by manipulating them to direct their trust badly. I suggest the epistemic individualist account should respond by adding a virtue of well-directed trust. The response’s success rests on a condition: well-directed trust is most plausibly cashed out in terms of an individualist virtue. In Chapter 2, I argue this condition fails because search engine bias exerts a crucial causal influence on how well a user directs their trust online. The failure leads me to conclude that epistemic structuralism is the most plausible account about explanation. But it is not necessarily the most plausible account about amelioration. One might focus primarily on improving users for pragmatic reasons. Yet, assuming epistemic structuralism about amelioration, I argue in Chapter 3 that reorganizing and relabeling information in an online environment might be more important improvement for information gathering than fixing flawed information gathering tools.

Table of Contents

| | |
|---|-----------|
| Introduction | 1 |
| 1: Analytic Framework: Epistemic Agents, Online Epistemic Tools and Online Environments | 4 |
| 2: Epistemic Individualism vs. Epistemic Structuralism | 7 |
| 3: Structure and Argument of the Thesis..... | 9 |
| Chapter 1: Is good epistemic character sufficient for using the Internet well? | 12 |
| 1: A Primer on Virtue Epistemology..... | 13 |
| 2: Heersmink’s Epistemic Individualist Virtue Responsibilism for the Internet (HIVR).... | 14 |
| 3: The Case of the Novice About Natural Medicine: A Problem for HIVR? | 17 |
| 3.2: Why the novice is virtuous in HIVR’s sense | 20 |
| 4: Echo Chambers: Two Challenges for Heersmink’s Epistemic Individualism..... | 23 |
| 4.1: A psychological model of testimony assessment online: background beliefs and trust | 24 |
| 4.2: Alternative explanation: the agent was led astray by rigged background beliefs | 26 |
| 4.3: Two epistemic structuralist challenges for Heersmink’s virtue responsibilism..... | 29 |
| 5: A Trustworthy Virtue Responsibilism for the Internet? | 30 |
| 5.1: Sketch of the virtue of well-directed trust into individualist virtue | 31 |
| 5.2: Addressing the explanatory challenge (EC)..... | 32 |
| 5.3: Addressing the practical challenge (PC) | 33 |
| Conclusion..... | 35 |
| Chapter 2: The virtue of well-directed trust: an individualist or extended virtue? | 36 |
| 1: Trust, Directing Trust, and the Epistemic Individualist Conception of the Virtue of Well-Directed Trust..... | 38 |
| 1.1: Trust as unquestioning attitude and its role in information gathering online | 38 |
| 1.2: The capacity to direct trust..... | 44 |

| | |
|--|------------|
| 1.3: What an epistemic individualist conception of the virtue of well-directed trust should look like..... | 47 |
| 2: The Virtue of Well-Directed Trust Is Most Plausibly an Extended Virtue..... | 50 |
| 3: The Virtue of Well-Directed Trust Not Extended, After All? | 55 |
| 3.1: Objection: Taylor was epistemically gullible for trusting the search engine | 55 |
| 3.2: Reply | 58 |
| Conclusion..... | 64 |
| Chapter 3: Biased online epistemic tools: flawed tools or flawed online environments? | 65 |
| 1: The Veritas Library | 67 |
| 2: Concepts for Flawed Epistemic Relevance Structure Arguments | 70 |
| 2.1: Epistemic relevance structure | 70 |
| 2.2: Epistemic order | 75 |
| 2.3: Conceptual difficulties with epistemic relevance structure and epistemic order | 77 |
| 3: Epistemically Disordered Environment as Structural Cause of Librarian’s Bias | 78 |
| 4: From the Veritas Library to Simplified Models of Facebook..... | 80 |
| 4.1: Simplified model of the Facebook information environment and its epistemic relevance structure..... | 81 |
| 4.2: Simplified model of the Facebook News Feed | 84 |
| 5: Facebook: Also an Epistemically Disordered Environment | 90 |
| 5.1: Inadequate epistemic information ordering principles..... | 91 |
| 5.2: Inadequate epistemic relevance cues..... | 93 |
| 5.3: Two structural causes of the Facebook News Feed’s bias | 96 |
| Conclusion..... | 98 |
| Conclusions and Reflections | 99 |
| Bibliography | 101 |

Introduction

To learn about the world, we ask questions about it. Our questions are varied: they range from “What is the antonym to ‘mundane?’” to “Is it safe to take the COVID-19 mRNA vaccine?”. To answer the questions we have, we must often gather relevant information.¹ Many of us turn to the Internet for this purpose. Roughly, the Internet is a system of various digital devices connected via a network allowing for communication of information.² Using various services, Internet users can gather information by, for instance, accessing web pages on the World-Wide Web or content posted by users on social media.³ On the face of it, this seems like good strategy: the Internet allows one to access a wealth of information on almost any topic, often free of charge.

However, the Internet face several challenges as a source of information. Under various names—misinformation, disinformation, fake news, to name a few—many have complained the Internet is a substantial source of *false* information.⁴ On this view, a central epistemic problem with the Internet is it contains a lot of false information that can make users form false beliefs.

We can all agree that forming false beliefs is an important epistemological problem. Yet, from the point of view of online information gathering, an exclusive focus on falsity is misplaced. Suppose the Internet only contained true information. Plausibly, Internet would still contain a lot of information on various topics. So, Internet users would still face the challenge of navigating the Internet to find the relevant piece of true information necessary to answer their question.⁵ The presence of false information in and of itself is not the only problem. Another important problem for online information gathering is the abundance of information

¹ As Jessie Munton has noted, acquiring information is a fundamental task for cognitively limited agents. Munton, "Perceptual Skill," 134.

² This definition is a paraphrase of *Store Norske Leksikon*'s entry. Øverby and Dvergsdal, "internett."

³ The examples of Internet services are from gathered Øverby and Dvergsdal's encyclopedia about the Internet. Øverby and Dvergsdal, "internett." Richard Heersmink also distinguishes between the Internet and the World Wide Web. Heersmink, "Virtue Epistemology," 5. Unless necessary, I will not refer to these services specifically. I am interested in the Internet as a whole, not just the particular services it provides.

⁴ For literature emphasizing the danger of misinformation and disinformation, see Zuboff, "Facebook and the Surveillance Society." For literature concerning “fake news,” see for example Croce and Piazza, "Consuming Fake News."; Nguyen, "Epistemic Bubbles."; Rini, "Fake News." For an account delineating the problems with all three, see Bergstrom and West, *Calling Bullshit*, 36-37.

Note that my sense of “information” is neutral. According to my neutral sense of the term, information need not necessarily be true—it can also be false. Thanks to Sebastian Watzl for urging me to clarify this term.

⁵ Thanks to Sebastian Watzl for making a structurally similar point about online environments, and for encouraging me to focus on epistemic relevance, as opposed to truth, throughout the thesis. My focus on relevance is also inspired by Sperber and Wilson's work on relevance theory. For an introduction to relevance theory, see Wilson and Sperber, "Relevance Theory."

epistemically irrelevant to an agent's question. Let us call this the "problem of epistemically irrelevant information." To be clear, I'm not saying false and epistemically irrelevant information are on par. Health beliefs based on false information seems to be more liable to lead to unhealthy decisions than health beliefs based on epistemically irrelevant information. My point is merely this: Insofar as a piece of information is epistemically irrelevant to answer an agent's question, the piece of information's truth or falsity is less important. So, the problem of epistemically irrelevant information is one of the central challenges for the Internet *qua* information source.

What is important for users to avoid epistemically irrelevant information online is not obvious, though. Many authors focus on users, i.e., epistemic individuals. To name a few, these authors claim users should have the right intellectual character traits, have a good method for finding experts online, employing critical thinking, and an ability to detect bullshit.⁶ On this view, provided Internet users have the right dispositions and conduct their information gathering online in the right way, the Internet is a good source of epistemically relevant information.

Therefore, one might be led to think individual factors are the primary factors making the Internet a source of epistemically relevant information. However, some authors in the epistemology literature have called the focus on individuals into question.⁷ Similarly, there are arguably good reasons to call into question a focus on individuals in information gathering online. Many factors outside the individual affect the quality of the Internet as a source of epistemically relevant information. For one, Internet users' online information gathering often rely on what I call "online epistemic tools." Roughly, these are tools internet users employ to acquire information online.⁸ A search engine like Google Search is a prime example of an online epistemic tool. Google Search orders web pages based on relevance. Online epistemic tools often rely on machine learning algorithms. Very roughly, machine learning algorithms

⁶ For focus on epistemic character, see Heersmink, "Virtue Epistemology." For an article providing a «method» for finding experts online, see Anderson, "Democracy." Bergstrom and West emphasize critical thinking and bullshit detection for dealing with misinformation. Bergstrom and West, *Calling Bullshit*, xii-xiv.

⁷ See for example Levy and Alfano, "Knowledge From Vice," 889-91. C. Thi Nguyen considers his account of trust as a step towards a "non-individualist epistemology." Nguyen, "Trust," 37.

⁸ The focus on tools is inspired by the extended mind literature's focus on tools as extending cognition. For an example of language as a tool for extended cognition, see Clark and Chalmers, "Extended Mind," 18. For an extended mind account focusing on Google Search, see Schwengerer, "Online Intellectual Virtues." However, as I state footnote 60 in Chapter 2, my account is not intended to be a metaphysical commitment to a version of the extended mind thesis.

solve problems by “learning” statistical patterns from data.⁹ One problem with machine learning algorithms is they can solve problems in biased ways by “learning” biases in the data.¹⁰ So, online epistemic tools relying on machine learning algorithms might be biased to present Internet users with epistemically irrelevant information. It is reasonable to assume the online epistemic tool algorithm’s bias is an important factor in making the Internet a good source of epistemically relevant information.

In addition, there is arguably yet another factor. Consider that online environments users and online epistemic tools are embedded in are structured in a particular way. Several authors have argued an environment’s structure have important epistemic consequences. For example, Jessie Munton has argued epistemically flawless agents embedded in an unjust social structure can “manifest” the biases of the social structure.¹¹ In a different context, Gabrielle Johnson, to give another example, has suggested machine learning algorithms can become biased by learning statistical patterns from data produced by an environment with a problematic social structure.¹² Similarly, I want to explore whether Internet users and online epistemic tools can manifest problems with how an online environment structures its information with respect to epistemic relevance.¹³

What are the most decisive factors for solving the problem of epistemically irrelevant information in online information gathering? And which factors should we focus on to make the Internet a *better* source of epistemically relevant information? These are important questions to address because many of us gather information online. We should also care about these questions because it can provide us with more focused interventions on making the Internet a better source of epistemically relevant information. Finally, these questions are interesting because online information gathering sheds light on offline information gathering. The Internet throws into sharp relief the relationship between epistemic agents, epistemic tools, and the environment they are embedded in.

⁹ Cf. Editors of Encyclopaedia Britannica, "algorithm." It is also based on the presentation of machine learning found in Bergstrom and West, *Calling Bullshit*, 182-83; Johnson, "Algorithmic bias," 9945-48.

¹⁰ For accounts on social biases in machine learning, see Geburu, "Race and Gender."; Johnson, "Algorithmic bias." In Chapter 3, footnote 1, I justify my focus on bias towards epistemic irrelevance instead social bias.

¹¹ For a concise formulation of this point, see Munton, "Perceptual Skill," 147.

¹² Johnson, "Algorithmic bias," 9948.

¹³ As chapter 3 will show, this idea is indebted to Jessie Munton’s notion of salience structure as organization around accessibility. Munton, "Prejudice," 1. Munton has herself started work on applying the salience structure concept to the Internet. See for example the abstract for her talk “How to Evaluate a Salience Structure” held at the GoodAttention Opening Workshop. GOODATTENTION, "GOODATTENTION: Opening Workshop."

In the following sections of this introductory chapter, I will delimit and restrict the scope of these questions. To this end, section 1 introduces my analytical framework, consisting of three components. Section 2 introduces the distinction between epistemic individualism and epistemic structuralism. Finally, section 3 provides the questions I try to answer in this thesis, as well as a summary of the structure and argument of the thesis.

1: Analytic Framework: Epistemic Agents, Online Epistemic Tools and Online Environments

The analytic framework I will use in this thesis consists of components: epistemic agents, online epistemic tools, and the online environment. Each of these have factors—*individual factors*, *epistemic tool factors*, and *environmental factors*, respectively—contributing to solving the problem of epistemically irrelevant information in online information gathering. Let us explain each component in turn.

The first component of the analysis is *epistemic agents*. By ‘epistemic agent’ I refer to individuals who turn to the Internet to gather epistemically relevant information about epistemic questions they have. For the rest of the thesis, unless stated otherwise, “agent” should be understood as “epistemic agent.” An example of an epistemic question an agent wants to find the answer to is: “How does the COVID-19 mRNA vaccine work?”. Answering epistemic questions constitute an epistemic goal of the agent.¹⁴

To answer their questions, agents conduct inquiries, where they gather epistemically relevant information online.¹⁵ From the point of view of the agent, online information gathering consists of two tasks: finding and assessing information online.¹⁶ There are numerous *individual factors* involved in accomplishing these two tasks. In my thesis, I will focus on three individual factors in particular: epistemic character, beliefs, and trust direction.

Roughly, ‘epistemic character’ refers to individual character traits disposing them to lesser or greater degree to find and assess epistemically relevant information online in a truth-conducive and error-minimizing manner.¹⁷ An agent with the epistemic virtue Richard Heersmink calls “intellectual tenacity,” for example, will be disposed to persist in finding and

¹⁴ I am following Richard Heersmink’s analysis of agents as motivated by achieving epistemic goals, see Heersmink, “Virtue Epistemology,” 3.

¹⁵ My focus on inquiry is inspired by a talk on inquiry by Susanna Siegel, which I attended at the GoodAttention Opening workshop at the University of Oslo in May 2022. GOODATTENTION, “GOODATTENTION: Opening Workshop.”

¹⁶ My conceptualization of the epistemic agent is influenced by Richard Heersmink’s emphasis on information finding and assessment online. Heersmink, “Virtue Epistemology,” 6, 9. (See also Chapter 1, footnote 24.)

¹⁷ This definition is indebted to Richard Heersmink’s notion of “cognitive character,” “intellectual virtues” and “intellectual vices.” Heersmink, “Virtue Epistemology,” 3-5. He defines virtues and vices explicitly in terms of being truth-conducive and error-minimizing.

assessing information online.¹⁸ Another important individual factor is the agent's beliefs. What the agent believes will be important for finding and assessing information. The final individual factor I want to highlight here is trust. Whom and what an agent decides to trust as a source of information is an important part of gathering information online. As chapters 1 and 2 will show, directing trust is associated with the agent's character and beliefs.

There are other individual factors which are relevant for information gathering online, though. For one, the agents in this thesis are for the most part novices. They do not have expertise in the domains they are gathering information about. For another, as I will focus on in Chapter 2, agents are cognitively and practically limited.¹⁹ Agents have limited cognitive resources to spend on finding and assessing information online. Also, there are only so many hours in a day an agent can realistically devote to information gathering online. As we will see in chapters 1 and 2, being a novice and having cognitive and practical limitations are important factors to consider when assessing what explains (un)successful gathering of epistemically information online.

The second component in my analytical framework is online epistemic tools. Roughly, "online epistemic tools" refer to tools agents use to aid their information gathering online.²⁰ Examples of online epistemic tools are search engines like Google Search and the Facebook News Feed. These online epistemic tools share a similar function: They order online information according to predicted relevance for the agent. On my simplified models of these online epistemic tools, they predict relevance for users by employing a machine learning algorithm.²¹ Insofar as online epistemic tools play an important role in online information gathering, the extent to which the online epistemic tool orders information according to epistemic relevance will be important. When an online epistemic tool orders information in a manner that presents the user with epistemically irrelevant information, I say the online epistemic tool's algorithm is "biased towards epistemic irrelevance". So, the *epistemic tool factor* of interest for this thesis is the bias towards epistemic relevance of the online epistemic tool's algorithm.

¹⁸ For Heersmink's definition of "intellectual tenacity," see Heersmink, "Virtue Epistemology," 4.

¹⁹ The focus on cognitive and practical limitation of agents is indebted to the work of Jessie Munton and C. Thi Nguyen. See especially Munton, "Perceptual Skill," 134; Nguyen, «Trust,, 45.

²⁰ Richard Heersmink uses the term "epistemic tool" in his discussion of using search engines virtuously. Heersmink, "Virtue Epistemology," 7. I call it "online epistemic tools" to underscore my thesis' focus on *online* information gathering.

²¹ For a rough definition of machine learning algorithm, see above.

The third component of my analysis is the online environment. Roughly, I understand this to be the surroundings in which the agent conducts their online information gathering, with or without online epistemic tools.²² As such, the Internet as a whole could be understood as an online environment. However, to reduce the scope of the thesis, my focus will be restricted to local online environments. Chapters 1 and 2 will consider the online environments one access via search engines. Chapter 3 will consider the Facebook online environment. Another measure to reduce the scope is to analyze online environments as information environments—more specifically as information environments consisting of written information. As we will see in chapter 3, I will consider the Facebook online environment as an information environment consisting of Facebook posts with written information. As the Internet consists of more than written information, this will be a limitation with my analytical framework. But it will be a helpful simplification to explore information gathering online. So, the thesis will consider local online environments understood as information environments with written information.

My thesis will focus on one *environmental factor*: epistemic structures.²³ I will use ‘epistemic structure’ in a broad sense and in a narrower sense. “Epistemic structure,” broadly construed, refers to the elements available in the online environment for the agent’s information gathering. One element in the broader epistemic relevance structure is the written information itself. The online environment needs to consist of written information the agent can gather to achieve their epistemic goals. This need not be the case. The information necessary to answer the agent’s question might not be present in the online environment. That said, throughout this thesis, I assume the online environment always has the information an agent needs to answer their question.

Information is also involved in a narrower sense of ‘epistemic structure,’ which will be the main focus of my thesis. As chapter 3 will argue, how information is structured for epistemic relevance in an information environment is an important aspect of information gathering. I will refer to this epistemic structure as the information environment’s “epistemic relevance structure.” This will be the primary focus of my thesis.

²² The definition of “environment” more broadly will be given and briefly discussed in chapter 3.

²³ For my focus on epistemic structures in particular, I am indebted to C. Thi Nguyen’s work, see Nguyen, “Cognitive islands.”; Nguyen, “Epistemic Bubbles.” Likewise, I am indebted to Croce and Piazza’s structural approach to fake news consumption online. Croce and Piazza, “Consuming Fake News.” I am influenced by Jessie Munton’s work on the influence of social structures. Munton, “Perceptual Skill.”; Munton, “Prejudice.” Finally, I am inspired by Gabrielle Johnson’s work on the influence of social structures on machine learning algorithms. Johnson, “Algorithmic bias.”

Other parts of the broader epistemic structure will also be discussed. One such element is the presence of online epistemic tools. Whether an online environment has online epistemic tools will often be crucial when the online environment is rich in written information. It would be harder for an agent to find a relevant webpage without the help of a search engine. Also, Chapter 1 considers the influence of what C. Thi Nguyen calls “social epistemic structures,” more specifically echo chambers.²⁴ As Chapter 1 argues, when an agent gathers information online using search engines, they can encounter web pages from online communities which qualify as an echo chamber. Insofar as agents can gather information from such online communities, they should be considered an element in the broader epistemic structure of an online environment.

Based on the previous discussion, I will categorize factors contributing to solving the problem of epistemically irrelevant information in online information gathering into two categories. The categories are individual factors and structural factors. The individual factors are simply the factors associated with the epistemic agents discussed above. Structural factors encompass the factors associated with the epistemic structure, such as the epistemic relevance structure. Additionally, since I consider the online epistemic tool a part of an online environment’s broader epistemic structure, I will also consider the epistemic tool factor a structural factor. That is to say, the bias towards epistemic relevance of an online epistemic tool’s algorithm is categorized as a structural factor.

In the next section, I will use the two categories of factors to construct two positions about solving the problem of epistemically irrelevant information in online information gathering.

2: Epistemic Individualism vs. Epistemic Structuralism

Using the distinction between individual and structural factors, we can distinguish between two positions with respect to the problem of epistemically irrelevant information in online information gathering. I call these positions “epistemic individualism” and “epistemic structuralism.”²⁵ These positions concern two topics: *explaining* the factors making agents

²⁴ Nguyen, “Epistemic Bubbles,” 142.

²⁵ My individualism-structuralism distinction is inspired by recent work in epistemology. Croce and Piazza distinguish between an “educational” and structural approach to improving fake news consumption on social media. Croce and Piazza, “Consuming Fake News,” 8. Relatedly, Bergstrom and West contrast what I take to be structural approaches to mitigating misinformation, disinformation and fake news—technological, governmental regulation—with educational approaches. Bergstrom and West, *Calling Bullshit*, 36-37. Similarly, Mark Alfano distinguishing between approaches aimed at improving individuals’ virtues and approaches focusing on fixing their testimonial network. For a brief summary, see Alfano, “Virtues for agents,” 8428-29.

My distinction differs insofar as I distinguish between an explanatory and amelioratory dimension. Also, my definition has a more general scope: it is meant to capture all individual factors (e.g., virtues), as well

gather epistemically relevant information online successfully or unsuccessfully, and *ameliorating* the online information gathering process. The distinction between epistemic individualism and epistemic structuralism is an important and helpful distinction. The distinction provides the ground for more specific explanations and interventions for making the Internet a better source of epistemically relevant information.

The first position I call “epistemic individualism.”²⁶ An epistemic individualist holds that when an agent finds epistemically relevant information online, this is *mainly* explained in terms of factors pertaining to the individual epistemic agent. For example, as I will discuss in Chapter 1, an epistemic individualist might claim successful online information gathering is primarily explained in terms of the user’s epistemic character. On this view, the user successfully finds epistemically relevant information online because they exercise epistemic virtues like thoroughness, open-mindedness, and so on. This is not to say that an epistemic individualist rejects the influence structural factors altogether. The epistemic individualist admits that certain structural preconditions must be met. They merely maintain that individual factors account for *most* of the explanation.

An epistemic structuralist about explanation, by contrast, maintains that when an agent finds epistemically relevant information online, this is *mainly* explained in terms of structural factors.²⁷ For example, as we will see in chapter 2, given agents’ cognitive and practical limitations, an epistemic structuralist could argue agents crucially depend on online epistemic tools to gather information online. If so, the bias towards epistemic irrelevance of the online epistemic tool’s algorithm will surely be decisive for finding epistemically relevant information online. Hence, the epistemic structuralist would argue, this shows epistemic structuralism to be the most reasonable position. Now, an epistemic structuralist will certainly

all types of structural factors (e.g., testimonial networks, how an online environment structures information with respect to epistemic relevance).

²⁶ “Epistemic individualism” (or similar observations capturing a primary focus on agents online and offline) has been employed by numerous authors. See for example Kawall, “Other-Regarding Epistemic Virtues,” 262; Nguyen, “Trust,” 37; Smart and Clowes, “Intellectual Virtues,” 9.

²⁷ My use of “epistemic structuralism” warrants two comments. First, epistemic structuralism should not be confused with the structuralist school of thought in sociology. While a comparison between epistemic structuralism and structuralism will likely be fruitful, I do not intend to draw upon that sociological literature in this thesis. Second, some philosophers have recently proposed a position known as “epistemic situationism.” Turri, Alfano, and Greco, “Virtue Epistemology.” To my understanding, epistemic situationism is meant to be a counterpart to virtue epistemology, just as situationism is a counterpart to virtue theory in ethics. There is overlap between epistemic situationism and epistemic structuralism. Yet, they should be considered as distinct position, insofar as epistemic structuralism emphasizes epistemic structures, which is not necessary to epistemic situationism. While I hypothesize epistemic structuralism might turn out to be a special instance of an epistemic situationist position, I do not have the space to pursue this hypothesis further in this thesis.

concede that individual factors play a role. However, the epistemic structuralist thinks the structural factors are most important.

Similarly, one might be an epistemic individualist or epistemic structuralist about amelioration.²⁸ That is to say, one can contend interventions on individuals or structures in the online environment will make the Internet a better source of epistemically relevant information. Note that epistemic structuralism about *explanation* does not necessarily imply commitment to epistemic structuralism about *amelioration*. One might hold a view that structural factors provide the best *explanation* of epistemic success or failure online. Yet, one might hold a view that individuals—their epistemic character, say—are a more feasible target for *improvement* to make the Internet a better source of epistemically relevant information.

3: Structure and Argument of the Thesis

With the analytical framework and the distinction between epistemic individualism and epistemic structuralism introduced, I turn to the main questions and the structure the thesis. In broad strokes, the question guiding the thesis is: Which account—epistemic individualism or epistemic structuralism—provides the best account of the factors that makes the Internet a good source of epistemically relevant information?

To concretize, I break this question into two different ones, reflecting the explanatory and practical distinction from the individualism-structuralism distinction:

The explanatory question: Which account—epistemic individualism or epistemic structuralism—provides the best *explanatory* account of the factors for epistemic success or failure in online information gathering?

The amelioratory question: Which account—epistemic individualism or epistemic structuralism—provides the best account of the factors that *improve* online information gathering?²⁹

Obviously, considering the entire Internet is beyond the scope of this thesis. Therefore, to answer the questions by focusing on search engines, as well as on the Facebook News Feed and the Facebook information environment.

²⁸ The structural approach to amelioration is indebted to similar structural proposed by Croce and Piazza and by Mark Alfano. Croce and Piazza, "Consuming Fake News," 8; Alfano, "Virtues for agents," 8428-29. Similarly, Smart and Clowes suggest reliability information in online systems might be more important than intellectual virtue for epistemic interaction with the Internet in certain contexts. Smart and Clowes, "Intellectual Virtues," 9. For more about how my definitions differ from these authors', see footnote 25 above.

²⁹ The ameliorative question is inspired by Alfano's questions for "ameliorative" social epistemology, focusing on improving agents or their networks. Alfano, "Virtues for agents," 8428-29. See also footnote 25 above.

The thesis explores these questions in three chapters. Each chapter has a different analytic component in focus. Chapter 1 focuses on epistemic agents and their contribution to online information gathering. To this end, I take Richard Heersmink's virtue epistemological account as a starting point.³⁰ On this view, by using search engines in an epistemically virtuous way, an agent can use the Internet as a good source of epistemically relevant information. I will interpret the account as an epistemic individualist account, both with respect to explanation and amelioration.

However, I argue that this individualist account underemphasize trust, a crucial phenomenon in online information gathering. Drawing upon work by C. Thi Nguyen, I raise an epistemic structuralist objection against this individualist account.³¹ An agent embedded in an echo chamber can be epistemically virtuous in Heersmink's sense yet fail epistemically relevant information and assess it as such. The echo chamber manipulates the agent to direct their trust towards bad experts. I argue that epistemic individualism can respond to the objection by adding a trust-related virtue, which I call "the virtue of well-directed trust". Yet, for the response to be successful, I argue it must meet a condition: the virtue of well-directed trust must be most plausibly conceptualized primarily in terms of the epistemic individual, i.e., as an individualist virtue.

In chapter 2, I assess this condition. To this end, the chapter brings in online epistemic tools more into focus. To assess whether the condition holds, I explore what trust is and what it means to trust well in an epistemic individualist sense. I draw on C. Thi Nguyen's understanding of trust as an unquestioning attitude, and I introduce a capacity to direct trust.³² I argue the epistemic individualist should conceptualize the virtue of well-directed trust as a disposition of the capacity to direct trust in truth-conducive and error-minimizing way.

However, I ultimately argue that the virtue of well-directed is most plausibly conceptualized as an extended virtue. That is, a virtue where at least one considerable cause of virtuous activity's result lies outside the epistemic agent. Roughly, I argue the virtue of well-directed trust is an extended virtue because a search engine's bias exerts a considerable causal influence upon whom cognitively and practically limited agents wind up trusting in their online information gathering. Thus, I conclude the virtue of well-directed trust is best

³⁰ Heersmink, "Virtue Epistemology."

³¹ Nguyen, "Epistemic Bubbles."; Nguyen, "Cognitive islands."

³² Nguyen, "Trust."

conceptualized as an extended virtue. Hence, the condition for the epistemic individualist response in Chapter 1 is not met.

In chapter 3, I will move our focus away from agents and towards online epistemic tools and online environments. Chapter 2 identified bias towards epistemic irrelevance as a central challenge for obtaining epistemically relevant information online. Chapter 3 investigates which structural improvements can mitigate this form of bias. A natural suggestion is that bias towards epistemic irrelevance is due to flawed online epistemic tools. Yet, I provide a different suggestion: bias towards epistemic irrelevance can also be “inherited” by an online environment structuring information badly with respect to epistemic relevance.

To assess these two hypotheses about bias, I turn my focus away from search engines and towards Facebook. The move from search engines is motivated primarily by the fact that Meta’s graph application programming interface (Graph API) provide us with a helpful model of how information is represented and stored in Facebook’s online environment.³³ I introduce simplified models of the Facebook information environment and the Facebook News Feed. I contend the Facebook information environment is not well-structured for epistemic relevance. As a result, I argue that, even if we were to optimize the Facebook News Feed’s algorithm for epistemic relevance, it would likely still order information in a biased manner. Thus, I conclude that changing an online environment’s epistemic relevance structure might, in some cases, be the most important structural factor for improving online information gathering.

In the concluding chapter, I summarize the result of the chapters. I contend that Chapter 2’s failure to plausibly establish the virtue of well-directed trust as an individualist virtue strongly suggests epistemic structuralism provides the best *explanatory* account. Yet, it does not necessarily mean epistemic structuralism is the best *amelioratory* account. Difficulties with making structural interventions could, for pragmatic reasons, make epistemic individualism the best amelioratory account. Yet, assuming epistemic structuralism were correct, Chapter 3 suggest that an online environment’s structure of information with respect to relevance is an important factor for improving online information gathering. I end with an exhortation for further research to focus on what the Internet is, and what we should want it to be.

³³ For an overview, see Meta, "Overview – Graph API." In the final stages of writing chapter 3, I discovered Google Search recently published a more detailed description. See for example Google, "In-Depth Guide to How Google Search Works." Future work could use this document to make a similar model of Google Search.

Chapter 1: Is good epistemic character sufficient for using the Internet well?

In the introductory chapter, I identified the problem of epistemically irrelevant information as an important problem for using the Internet as a source of information. In this chapter, I want to consider an account by Richard Heersmink implying the onus for solving the problem is primarily on users.¹ Heersmink argues epistemic virtues—e.g., curiosity, thoroughness, and open-mindedness—make it highly probable users will use the Internet well by finding and assessing information using search engines. There are several reasons why I want to focus on this account. First, as I will argue in this chapter, Heersmink's emphasis upon epistemic character makes it a good example of an epistemic individualist position. On my interpretation, the account implies Internet users can solve the problem of epistemically irrelevant information by having good epistemic character traits. Second, the account prominently features online epistemic tools in online information gathering. The use of online epistemic tools is an important aspect of online information gathering. Also, as I will argue in Chapter 2, the role of online epistemic tools in online information gathering call epistemic individualism into question.

However, Heersmink's account does not emphasize trust in online information gathering. C. Thi Nguyen argues echo chambers, a social epistemic structure, manipulate its members trust.² On this view, trust manipulation can make even a good epistemic agent produce epistemically bad results. If so, this suggests the social epistemic structure the agent is embedded in is more decisive than their epistemic character for their epistemic success or failure online. That is, it suggests good epistemic character traits are not sufficient for using the Internet well. The question I want to address in this chapter is: Are good character traits sufficient for making Internet users solve the epistemically irrelevant information problem?

The structure of the chapter is as follows. Section 1 will give a general introduction to virtue epistemology. This lays the foundation for Heersmink's individualist virtue responsibilist account (HIVR, for short), which I will present in section 2. Section 3 presents the case of Karl. Karl is meant to be an example of an agent, virtuous in HIVR's sense, yet whose search engine behavior produces epistemically undesirable results. Section 4 presents an epistemic structuralist alternative explanation, arguing the Karl case is best explained as trust-manipulation of an echo chamber. I will argue this epistemic structuralist argument presents both an explanatory and practical problem for HIVR. Section 5 argues HIVR can respond by

¹ Heersmink, "Virtue Epistemology."

² Nguyen, "Epistemic Bubbles."

adding a virtue of well-directed trust (“WDT”, for short). I call this the “augmented individualist virtue responsibilism” (“AIVR,” for short). The thesis of this chapter is that AIVR can *in principle* respond to and meet the explanatory and practical challenge. Yet, the response rests on a crucial condition that trusting well is most plausibly accounted for as an epistemic virtue (or at least primarily in terms of factors pertaining to the epistemic individual).

1: A Primer on Virtue Epistemology

To better understand Heersmink’s virtue epistemological account for the Internet, which I will present in section 2, this section aims to provide a background on virtue epistemology (VE).

I want to highlight two features of VE theories which, according to John Turri, Mark Alfano, and John Greco, distinguish VE theories from other epistemological theories.³ First, the agent and groups are objects of evaluation.⁴ VE theories focus on cognitive properties—e.g., epistemic character—rather than beliefs.⁵ Instead of evaluating an agent based on the beliefs they hold—whether the agent’s beliefs are justified and true, say—VE theories tend to evaluate an agent based on their epistemic character traits.⁶ Second, VE theories have a normative focus.⁷ According to Turri, Alfano, and Greco, this means VE theories generally focus on ‘[...] epistemic norms, values and evaluation’.⁸ In this way, VE separates itself from epistemological theories that oppose taking a stance on what is reasonable to believe.⁹ VE theories have a substantial normative component: they hold epistemology should help agents cultivate intellectual flourishing.¹⁰

To illustrate the two features of VE theories, consider an agent who has come to believe that Joe Biden is the 46th president of the United States. Starting with the first feature, some epistemological theories focus less on the agent and more on the *belief* itself. For example, whether there are good grounds for holding the belief, and whether the belief is in fact true. By contrast, VE theories focus on the *agent*. For example, whether the agent came to hold the

³ Turri, Alfano, and Greco, "Virtue Epistemology."

⁴ Turri, Alfano, and Greco, "Virtue Epistemology." The authors note that some VE theories also focus on evaluating the character traits of epistemic communities. The focus in this chapter, however, are agent-centered strands of virtue epistemology.

⁵ Heersmink, "Virtue Epistemology," 2; Turri, Alfano, and Greco, "Virtue Epistemology." But see Nancy Daukas’ virtue epistemology, where beliefs play a more prominent role. Daukas, "Epistemic Trust."; Daukas, "Altogether Now."

⁶ Heersmink, "Virtue Epistemology," 2.

⁷ Heersmink, "Virtue Epistemology," 2-3; Turri, Alfano, and Greco, "Virtue Epistemology."

⁸ Turri, Alfano, and Greco, "Virtue Epistemology."

⁹ Turri, Alfano, and Greco, "Virtue Epistemology."

¹⁰ Cf. Heersmink, "Virtue Epistemology," 2-3; cf. Turri, Alfano, and Greco, "Virtue Epistemology."

belief about Biden as a result of good character traits. This brings in the normative component, the second feature. If the agent came to hold the belief about Biden as a result of good intellectual traits, the agent receives a positive evaluation. Conversely, if the belief results from bad character traits, the agent receives a negative evaluation. To my understanding, according to VE theories, an agent can still receive a positive evaluation, even if the good intellectual traits produce false beliefs.¹¹

Based on their conception of the nature of good character traits (i.e., epistemic virtues), VE is ordinarily divided into two subgroups.¹² The first subgroup, virtue responsibilism, conceptualize epistemic virtues as “faculty-virtues”: cognitive faculties that reliably produce true beliefs.¹³ When it reliably productive of true perceptual beliefs, hearing is an example of a faculty-virtue. When it does not, it is a faculty-vice. Virtue responsibilism, the second subgroup, conceptualize epistemic virtues as “trait-virtues”.¹⁴ Heersmink’s view characterize trait-virtues as “[...] character traits that a person who desires truth and understanding would want to have.”¹⁵ Heersmink gives open-mindedness as an example of trait-virtue, and dogmatism as its corresponding trait-vice.¹⁶

With this primer on virtue epistemology, we can turn to Heersmink’s *virtue responsibilist* account. ‘Epistemic virtue’ (or simply ‘virtue’) henceforth refers to trait-virtue.

2: Heersmink’s Epistemic Individualist Virtue Responsibilism for the Internet (HIVR)

This section aims to provide an overview of Richard Heersmink’s argument for how good character traits can help agents use search engines in a truth-conducive manner that minimize error.¹⁷ As this account puts primary explanatory and practical emphasis upon agents and their epistemic character, I will argue he is a good representative of an epistemic individualist account.¹⁸ This suggests that good character traits can help agents solve the problem of epistemically irrelevant information online.

¹¹ Cf. Heersmink, "Virtue Epistemology," 2.

¹² Heersmink, "Virtue Epistemology," 2. Turri, Alfano and Greco note that some find the division unproductive, as both virtue types are necessary parts of epistemology. Turri, Alfano, and Greco, "Virtue Epistemology." As HIVR is a virtue responsibilist account, I will not pursue this criticism further.

¹³ Heersmink, "Virtue Epistemology," 2; Turri, Alfano, and Greco, "Virtue Epistemology."

¹⁴ Heersmink, "Virtue Epistemology," 2.

¹⁵ Heersmink, "Virtue Epistemology," 4.

¹⁶ Heersmink, "Virtue Epistemology," 4-5.

¹⁷ Heersmink, "Virtue Epistemology."

¹⁸ What I call “epistemic character” and “epistemic virtue” Heersmink calls “cognitive character” and “intellectual virtue”, respectively. Cf. Heersmink, "Virtue Epistemology," 2, 3.

I take Heersmink's main claim to be that good character traits (i.e., epistemic virtues) makes an agent use search engines in an epistemically good way. He considers the World Wide Web (or simply 'the Web') as an information space.¹⁹ Roughly, the Web refers to the interconnected webpages on the Internet.²⁰ One way agents find information in this space is by using search engines.²¹ How well an agent uses search engines depends on how virtuous their character is. This follows from Heersmink's definition of virtue and vice. He defines virtues as '[...] cognitive character traits that are truth-conducive and minimalise error'.²² By contrast, vices are '[...] cognitive character traits that are *not* [my italics] truth-conducive and do *not* [my italics] minimalise error'.²³ A virtuous agent will use search engines in a truth-conducive manner that minimize error. As I understand Heersmink, they do so by search engines to seek and assess information well (henceforth "search engine behavior").²⁴ While he notes his the list of virtues he discusses are not exhaustive, I take his central claim to be that, together, they are sufficient for using search engines well.²⁵ Thus, an agent with the virtues discussed by Heersmink will navigate the Web well by using search engines to find and assess information. I understand this to be the reason why his account strongly emphasizes the role of education in helping agents cultivate and practice virtues.²⁶ To concretize the argument and introduce the virtues, let us adapt one of Heersmink's own examples.²⁷

Suppose an epistemically virtuous agent, Clara, wants to learn about the pros and cons of COVID-19 vaccines. Clara starts by submitting a search query: "What are the pros and cons of COVID-19 vaccines?". Clara will be *curious* to explore a variety of the search results.²⁸ She will evaluate the results with *intellectual autonomy*. That is, she "thinks for herself",

¹⁹ Heersmink, "Virtue Epistemology," 5.

²⁰ Heersmink, "Virtue Epistemology," 5. More precisely, Heersmink understands the Web as an 'application' or 'layer' of the Internet. The technical details of these terms are not crucial for my argument.

²¹ Heersmink, "Virtue Epistemology," 1.

²² Heersmink, "Virtue Epistemology," 3.

²³ Heersmink, "Virtue Epistemology," 4.

²⁴ Heersmink does not formulate this explicitly, but I maintain it follows from his account. I submit the following quotes as evidence. He writes: cultivation of virtues is "[...] beneficial for one's *information-seeking behaviours* [my italics] [...]." Heersmink, "Virtue Epistemology," 9. Elsewhere, he writes that he wants to investigate "[...] how virtue epistemology can help Internet users to evaluate and determine which online information is epistemically useful and epistemically useless." Heersmink, "Virtue Epistemology," 6. The overall implication, I take it, is epistemic virtues will help agents *both* seek and assess information online in a good way.

²⁵ For the list of epistemic virtues, see Heersmink, "Virtue Epistemology," 3-4. For the claim that they are not exhaustive, see Heersmink, "Virtue Epistemology," 4.

²⁶ Heersmink, "Virtue Epistemology," 9-10.

²⁷ Heersmink, "Virtue Epistemology," 6-8. My adapted example follows Heersmink's presentation in introducing the virtues as part of an example.

²⁸ Compare to Heersmink's definition and description of curiosity. Cf. Heersmink, "Virtue Epistemology," 3, 7.

assessing search results with a skeptical attitude and without relying too much on others.²⁹ However, when appropriate, Clara defers to epistemic authority, for example to doctors.³⁰ Clara is also *intellectually humble*, aware of the limits both of her own capacities and of search engines.³¹ For instance, she is aware search engines can provide misleading suggestions, as well as presenting search results based on her earlier search queries.³² Furthermore, Clara seeks and assess information on search engines with *attentiveness*, *intellectual thoroughness*, and *intellectual carefulness*.³³ She does her utmost to pay attention to answer her question, seeking a comprehensive account about COVID-19 vaccines, assessing search results carefully and critically to avoid forming false beliefs. Finally, Clara engages with search engines with *intellectual tenacity*, *intellectual courage*, and *open-mindedness*.³⁴ She persists to find and assess a wide array of search results about COVID-19 vaccines (even if encountering disagreement causes discomfort) and revise her beliefs if a different view seems more accurate. By exercising her virtues, HIVR suggests Clara will find and assess information that will likely lead to knowledge and understanding about COVID-19 vaccines.³⁵

In light of the argument about the agent being the decisive factor in using search engines well, we can better understand Heersmink claim that “[w]e have a responsibility to use the Internet in an epistemically virtuous way.”³⁶ However, agents have different degrees of responsibility to use the Internet. He argues the agent’s responsibility is curtailed by “epistemic practicability.”³⁷ As Heersmink puts it, “[...] what an agent is required to do to obtain true beliefs thus depends on one’s cognitive capacities and contextual aspects.”³⁸ For example, a

²⁹ Compare Heersmink’s definition and description of intellectual autonomy. Cf. Heersmink, "Virtue Epistemology," 3, 7.

³⁰ Cf. Heersmink, "Virtue Epistemology," 11n5.

³¹ Compare to Heersmink’s definition and description of intellectual humility. Cf. Heersmink, "Virtue Epistemology," 3, 7.

³² Cf. Heersmink, "Virtue Epistemology," 7.

³³ Compare to Heersmink’s definition and description of attentiveness, intellectual thoroughness, and intellectual carefulness. Cf. Heersmink, "Virtue Epistemology," 4, 7.

³⁴ Compare to Heersmink’s definition and description of intellectual tenacity, intellectual courage and open-mindedness. Cf. Heersmink, "Virtue Epistemology," 4, 7-8.

³⁵ Compare this with Heersmink’s claim that virtues increase likelihood of obtaining knowledge. Cf. Heersmink, "Virtue Epistemology," 2.

³⁶ Heersmink, "Virtue Epistemology," 8. Two remarks are in place. First, Heersmink notes search engines and their creators bear epistemic responsibility, as well. Heersmink, "Virtue Epistemology," 8. However, he does not make clear how much responsibility he thinks they bear. Second, Heersmink’s account does not discuss the extent to which individuals are *responsible* for cultivating a good epistemic character. Plausibly, there will be an epistemic practicability component to acquire good character too: they will only be responsible insofar as it is epistemically practicable for them to cultivate it.

³⁷ Heersmink, "Virtue Epistemology," 8.

³⁸ Heersmink, "Virtue Epistemology," 8.

person living in a country with limited Internet access is not as epistemically responsible as someone who has widespread Internet access. Nonetheless, he retorts, these aspects serve as good reasons for the agent to cultivate a good epistemic character.³⁹

Because Heersmink puts such a strong emphasis upon epistemic character in his discussion of search engines, I consider him a representative of the position I called ‘epistemic individualism’ in the introductory chapter, both in the explanatory and the amelioratory sense.⁴⁰ Recall that epistemic individualism primarily emphasizes agents their explanations or suggestions for improving information gathering online. This fits well with the virtue responsibilist account we have considered in this section. Plausibly, Heersmink’s account suggests epistemic character explains whether the agent will use search engines well or badly. Similarly, if we want agents to use search engines in a better way, Heersmink’s account seems to me to suggest the agent must cultivate their epistemic character. Thus, it seems reasonable to think of Heersmink’s virtue responsibilism as an epistemic individualist account in both the explanatory and practical sense. To underscore its individualist aspects, I have chosen to call the account ‘Heersmink’s individualist virtue responsibilism’ (“HIVR,” for short).

3: The Case of the Novice About Natural Medicine: A Problem for HIVR?

In the last section, we saw HIVR claims that possessing the virtues he listed is sufficient for using search engines in a good way. This section aims to provide a case that challenges this claim. Heersmink remarks in passing that “naïve” agents are vulnerable to be led astray, for example about vaccines, when using search engines.⁴¹ I want to bring out that novices create bigger problems for HIVR than he seems to realize. In section 3.1, I will present a case where a novice agent, virtuous in HIVR’s sense, does not use search engines in a good way. Section 3.2 argues the agent is indeed virtuous in HIVR’s sense. I will argue that, although there might be something correct about the agent not having the virtue of open-mindedness, he nonetheless satisfies the definition of open-mindedness employed by Heersmink. In section 4, I argue this case highlights problems with HIVR, making it vulnerable to epistemic structuralist objections.

³⁹ Heersmink, "Virtue Epistemology," 8.

⁴⁰ Smart and Clowes also note the individualism in epistemic virtue responsibilist accounts about the Internet. Smart and Clowes, "Intellectual Virtues," 9.

⁴¹ Heersmink, "Virtue Epistemology," 6.

3.1: The case of the novice about natural medicine

Let us consider Karl, an agent slightly different from the epistemically virtuous Clara.⁴² Karl was raised in a community sympathetic to “natural medicine”. The community maintains that illnesses are best prevented and treated by using “natural” remedies: plant-based products, various forms of meditation and by avoiding “toxin”-rich foods. As a child, Karl acquired these beliefs by trusting epistemic authorities in the community—his parents and teachers, among others. A child is plausibly epistemically blameless for acquiring beliefs in this manner.⁴³ Aside from a curriculum with a penchant towards natural medicine, the educational system helps their pupils and students cultivate good epistemic character. Moreover, as a result, he reflects upon his beliefs, finding he has good evidence for them. People in the community, including himself, are in excellent shape and seldomly ill. As I will argue in section 3.2, Karl has a good epistemic character. The difference between Karl and Clara is Karl’s doxastic sympathy towards natural medicine.

Suppose the COVID-19 pandemic hits, and Karl must decide whether to take the COVID-19 vaccine.⁴⁴ To decide, he seeks a comprehensive explanation and understanding of COVID-vaccines. Like Clara, he submits a search engine query: “What are the pros and cons of the COVID-19 vaccine?”. The enquiry produces millions of search results. To find the most epistemically relevant information, he takes two measures. First, he uses the following learning principle: Consider search results by trustworthy experts on the topic of health and vaccination. Second, knowing the limitations of search engines, he considers multiple search results pages to find relevant expert search results.

During his search, he finds and considers a search result from a natural medicine expert. Because of his upbringing, he considers people with expertise in natural medicine trustworthy experts about health. The expert writes COVID-19 vaccines are dangerous, and natural

⁴² Three remarks are in place. First, the case is based on C. Thi Nguyen’s fictional case of a teenager raised in an echo chamber. Nguyen, "Epistemic Bubbles," 154-55. I have adapted it to the context of search engines. Thanks also to Sebastian Watzl for helpful suggestions. Second, the scenario is admittedly idealized in numerous respects. Most important is the time necessary to be led astray. In real-life cases, agents probably need to be exposed to more content, over longer periods of time, than the agent in my case does. Nonetheless, my conviction is the case captures the essence of being led astray by one’s flawed background beliefs. Third and last, my choice of natural medicine as knowledge domain is not intended to disparage supporters of natural medicine. People can find utility and value in it. And people can distrust in traditional medicine in favor of natural medicine for understandable reasons (e.g., being member of a group historically subjected to unethical medical experiments). Thanks to Ane Maria Gerdes Døhl for discussion on this topic.

⁴³ Nguyen contends children are plausibly epistemically blameless for trusting seeming epistemic authorities. Nguyen, "Epistemic Bubbles," 154. By extension, I assume the child would be blameless for any beliefs acquired via these trusting relationships.

⁴⁴ A thorough treatment of the case should add the complexity of multiple COVID-19 vaccines. My argument does not hinge on the existence of a variety of vaccines. So, for simplicity, I assume there is only one vaccine.

medicine alternatives are safer.⁴⁵ He cites scientific-sounding evidence: the vaccine lacks animal studies, and side effects are serious. To substantiate his claims, he refers to other experts. Karl diligently scrutinizes and follows the trail of sources to other expert web sites to verify the claims of the first expert. One of the natural medicine experts connects COVID-19 vaccination to a plot to depopulate the world.⁴⁶ He claims that he (and people who share his beliefs) know the “real” truth about COVID-19 vaccination. He states: Vaccination is part of a plot to depopulate the world. Vaccine advocates refuse to talk about this, he adds. When confronted, he says, they will deny their involvement and the dangers of COVID-19 vaccines. This is because they are trying to hide the truth, or because they are brainwashed.

Karl has found a network of (to him) trustworthy experts who have provided ample evidence and arguments about the dangers of COVID-19. However, although it makes him uncomfortable, he knows he should nonetheless also consider sources that disagree. He decides to assess the information WHO’s web page about the COVID-19 vaccines. He finds that WHO are seemingly vindicating what the natural medicine expert said. They strike him as underestimating the dangers associated with the vaccine. And they strike him as actively working to persuade people to take the vaccine by “debunking” vaccine myths about vaccines. Karl finds this suspicious. It seems to him WHO is hiding something. As a result, not only is Karl unconvinced by the arguments and evidence by the WHO, he also finds WHO a distrustful source about vaccines.⁴⁷ The natural medicine experts were right, he concludes.

Clearly, something has gone epistemically wrong. Karl has a good epistemic character in HIVR’s sense. But the result of exercising his virtues was that he acquired false beliefs about COVID-19 vaccines and came to distrust WHO as a source of information of vaccine. If this is correct, this calls into HIVR’s claim about epistemic virtues helping agents use search engines to acquire knowledge and deal with the problem of epistemically irrelevant information.⁴⁸ An obvious response to this case is to insist Karl is surely epistemically vicious in some way. In the next section, I argue that Karl in fact *does* possess the virtues of HIVR.

⁴⁵ I’m not claiming natural medicine *necessarily* implies hostility to vaccines. See also objection in footnote 56)

⁴⁶ The natural medicine expert’s argumentation is inspired by echo chambers’ disagreement reinforcement mechanisms in Nguyen’s analysis. Nguyen, “Epistemic Bubbles,” 147-48. I return to these in section 4.

⁴⁷ Lorenz-Spreen et al.’s systematic review found a negative association between digital media, institutional trust, and vaccines. Lorenz-Spreen et al., “Digital Media and Democracy,” 10. Thus, Karl’s distrust in the WHO has some empirical plausibility. Yet, the authors add, some evidence suggest a positive association between digital media and trust in science and in democracy. A thorough treatment of the empirical relationship between digital media, trust, and science is beyond the scope of this chapter and thesis.

⁴⁸ Recall Heersmink underscores epistemic virtues do not *guarantee*, but greatly *increase the likelihood* of obtaining knowledge. Heersmink, “Virtue Epistemology,” 2.

3.2: Why the novice is virtuous in HIVR's sense

This section aims to argue Karl is epistemically virtuous according to HIVR. I will start with what I take to be the least controversial virtues, before considering the more controversial ones. Open-mindedness has pride of place in my discussion because it is the most controversial one.

Starting with the less controversial virtues, Karl possesses the virtues curiosity, attentiveness, intellectual thoroughness, and intellectual courage. He is *curious* because he is driven to learn about the vaccines, both from a practical standpoint but also for its own theoretical sake.⁴⁹ He does this in an *attentive* way, focusing on the information most relevant to his epistemic enquiry.⁵⁰ The agent is *intellectually thorough* insofar as he seeks a comprehensive explanation of COVID-19 vaccines, and does not settle for a superficial account.⁵¹ And he is *intellectually courageous* because, despite finding it uncomfortable, he engages with the WHO's claims, which differ from his natural medicine beliefs.⁵²

Let us move on to the more controversial virtues: intellectual carefulness, intellectual tenacity, intellectual humility, intellectual autonomy and open-mindedness. He is *intellectually careful* because he takes measures to avoid forming false beliefs.⁵³ Based on his initial beliefs, he scrutinizes sources in the search results, evaluating their claims and evidence using logic and critical thinking.⁵⁴ That is, he is careful to reason using valid and sound rules of inference, and he is careful to assess claims and evidence skeptically before he accepts them.

Karl is *intellectually tenacious* and *intellectually humble*. Despite the difficulties of finding good information with millions of personalized search results, he nonetheless persistently seeks a comprehensive account of the COVID-19 vaccine.⁵⁵ At least, he is as tenacious as we can reasonably expect him to be.⁵⁶ Karl is also intellectually humble. Put in the terms of this

⁴⁹ Cf. Heersmink, "Virtue Epistemology," 3, 7.

⁵⁰ Cf. Heersmink, "Virtue Epistemology," 4, 7.

⁵¹ Cf. Heersmink, "Virtue Epistemology," 4, 7.

⁵² Cf. Heersmink, "Virtue Epistemology," 4, 7-8.

⁵³ Cf. Heersmink, "Virtue Epistemology," 4, 7-8.

⁵⁴ Cf. Heersmink, "Virtue Epistemology," 4, 7.

⁵⁵ Cf. Heersmink, "Virtue Epistemology," 4, 8.

⁵⁶ It could be objected that surely pro-vaccine websites by natural medicine experts are among Karl's search results. If Karl failed to find them, the objection goes, Karl is not virtuous because he is not tenacious enough.

My response is two-fold. First, the objection presupposes a tenacious search uncovering pro-vaccine natural medicine experts is epistemically practicable. With millions of search results, it is certainly epistemically impracticable to go through them all. In a worst-case scenario, it might not be epistemically practicable for the agent to find pro-vaccine natural medicine experts. Hence, as we have seen, they are not epistemically responsible for doing so, according to Heersmink's analysis. Note that in such a worst-case scenario, the agent can still have acted epistemically virtuously. To put HIVR to the ultimate test on search engines, I am assuming Karl finds himself in this worst-case scenario. Second, even if Karl were to identify vaccine-friendly expert, they

chapter, Heersmink understands intellectual humility as an awareness and admission of one's limitations in three respects: one's intellectual skills, knowledge and the online epistemic tools one uses (e.g., search engines).⁵⁷ Because he is aware and admits his lacking skills and knowledge pertaining to vaccines, Karl has decided to defer to (what he takes to be) trustworthy experts. Moreover, as we saw, he is aware and take measures to compensate for the limitations of search engines. For these reasons, it is reasonable to think Karl is intellectually humble.

Karl is also *intellectually autonomous*. According to the definition employed by Heersmink, intellectual autonomy consists of '[...] a willingness and ability to think for oneself,' while at the same time being '[...] realistic about one's cognitive skills and accept authority and experts when appropriate'.⁵⁸ Karl exhibits both aspects of intellectual autonomy. He thinks for himself insofar as he independently scrutinizes the claims and evidence of sources. Yet, knowing he is not an expert on vaccines, he willingly defers to (those he considers) trustworthy experts on vaccines. Thus, Karl is also intellectually autonomous.

Lastly, and most controversially, Karl is *open-minded* in Heersmink's sense. His definition of open-mindedness has two components. First, the open-minded agent has '[...] an awareness of which domains of knowledge [she is] knowledgeable about and of which domains [she is] less knowledgeable about'.⁵⁹ And second, she is '[...] willing to consider alternative views, and if these *seem* [italics mine] more accurate or well founded, then she is willing to change and revise her initial beliefs'—especially in knowledge domains she is less knowledgeable about.⁶⁰ I will argue Karl satisfies both of these.

Starting with the first component, it is plausible to think Karl satisfy the awareness component open-mindedness. It is unclear how comprehensive HIVR think this awareness should be. A complete, comprehensive overview over knowledge domains is surely a very high bar to demand from agents. At any rate, we saw Karl is aware he is not knowledgeable about vaccines. That is why he conducts the search query and experts the first place. So, at least in the context of COVID-19 vaccines, he satisfies the awareness component. As for the second

would face the difficulty of disagreeing experts. In that case, a skeptic could say, the agent ought to suspend their belief in effectiveness of vaccines. From a purely theoretical point of view, I grant this is a rational response. However, the serious consequences of the pandemic forces the agent to make up their mind. It is not obvious to me Karl will necessarily choose to believe the vaccine-friendly natural medicine expert.

⁵⁷ Cf. Heersmink, "Virtue Epistemology," 3, 7.

⁵⁸ Heersmink, "Virtue Epistemology," 3.

⁵⁹ Heersmink, "Virtue Epistemology," 4.

⁶⁰ Heersmink, "Virtue Epistemology," 4.

component, he was willing to consider alternative views. As we saw, he *did* evaluate the information provided by WHO. Although most of us would maintain WHO has more accurate and well-founded beliefs than the natural medicine experts, this was not how it *seemed* to Karl. This is in keeping with the second component of open-mindedness. It only requires agents to change their mind if the view *seems* more accurate and well-founded. Hence, according to HIVR's definition of open-mindedness, Karl is open-minded.

One could object to my claim that the agent satisfies the second component. Karl did not *truly* engage with the argument and evidence from WHO.⁶¹ He quickly deemed them to be untrustworthy. Because he considered them untrustworthy, the objector concludes, the agent is not truly open-minded. In reply, I want to highlight a crucial distinction: (a) being *in principle* disposed to reject beliefs simply because they differ from one's own, and (b) rejecting beliefs other than one's own *on the grounds of the source being untrustworthy*. Rejecting—at least suspending belief in—a line of argument because the source is untrustworthy (b) is compatible with being open-minded (a). To see why, consider the following example. We might in principle willing to listen to other views about nutrition and change our mind about it. Yet if we are listening to a nutrition expert with a track-record of being involved in fraudulent studies, we will at least our suspend belief until the claims are corroborated from more trustworthy sources.⁶² Note that this does not imply closed-mindedness. We are still in principle open to changing our mind about nutrition, even if we are suspicious of the expert. Similarly, Karl can also be open-minded in this sense, while being suspicious of WHO. Certainly, his belief about the WHO's involvement in a conspiracy is mistaken. Yet, *given* this belief, Karl does not necessarily lack open-mindedness. He might merely consider them (mistakenly) as an untrustworthy source of information about vaccines.

The objector could insist that there is an important difference between the rational suspicion associated with open-mindedness and the distrust displayed by Karl. Karl has come to believe that *all* vaccine supporters are in some way part of a conspiracy. But then Karl is effectively unwilling to change his beliefs about vaccination. Hence, the objector concludes, there is scant difference between categorical rejection of sources and lacking open-mindedness. I agree that they are very similar in this case, but lacking open-mindedness and problematic trust strikes me as distinct phenomena.⁶³ Moreover, as we will see in section 4, Karl's

⁶¹ For a discussion of the same point in the context of work by C. Thi Nguyen, see footnote 91.

⁶² For more on the utility of an expert's past track-records for novices, see Goldman, "Experts," 106-08.

⁶³ This conclusion warrants two comments. First, a distinction between open-mindedness and trust-patterns also seem to be present in Nguyen's paraphrase of an account of an echo chamber escapee: "He had to completely

problematic activity need not be explained as a problem of epistemic character. It can be explained by the social epistemic structure he is situated in. For these reasons, I claim Karl is open-minded—at least in HIVR’s sense.

If my line of reasoning in this section has been correct, Karl possesses every virtue of HIVR, but does not engage with search engines in a good way. This calls into question the claim that good epistemic character traits are sufficient for using search engines well, thereby dealing with the problem of epistemically irrelevant information. In the next section, I will provide a different explanation for the case of Karl, one where his predicament can be explained as the influence of social epistemic structures.

4: Echo Chambers: Two Challenges for Heersmink’s Epistemic Individualism

In this section, I want to provide an explanation of what happened to Karl without invoking Karl’s epistemic character. Bringing in work by C. Thi Nguyen, I will argue Karl’s epistemic behavior can be explained as the influence of an echo chamber, a possibility HIVR does not address.⁶⁴ An echo chamber is an example of what Nguyen calls an “social epistemic structure.”⁶⁵ Nguyen does not provide an explicit definition of “social epistemic structure.” Roughly speaking, by “social epistemic structure,” I will understand an epistemic network where information is exchanged between sources and receivers of testimony.⁶⁶ In the scenario above, Karl is situated in an epistemic network where he receives testimony about COVID-19 from the natural medicine experts via their webpages.⁶⁷ I will argue Karl’s behavior can be explained in terms of his epistemic network manipulating his trust, making it an echo chamber. To understand the influence of echo chambers, I will first present a model of testimonial assessment in section 4.1. With a grasp on this model, section 4.2 presents an overview of Nguyen’s analysis of how echo chambers manipulate the trust of their

abandon his belief system, and he spent years re-building a world-view of his own, immersing himself broadly and *open-mindedly* [my italics] in everything he’d missed [...] all with *an overall attitude of trust* [my italics].” Nguyen, “Epistemic Bubbles,” 158. This supports a view of open-mindedness as related, but distinct from trust.

Second, in principle one could perform a test to distinguish the two. Karl could be exposed to the WHO’s argument from a source he trusts. If he still rejects the arguments, then Karl *does* lack open-mindedness. If he does not, then the problem is Karl’s trust. But this test will likely fail in practice. Karl could simply interpret the heretofore trustworthy source as having been corrupted. Open-mindedness, trust, and intellectual autonomy) are likely related. The precise relationship between the three are beyond the scope of this chapter.

⁶⁴ Nguyen, “Epistemic Bubbles.”; Nguyen, “Cognitive islands.”

⁶⁵ Nguyen, “Epistemic Bubbles,” 142.

⁶⁶ This network definition is inspired by the terms “epistemic network” and “testimonial network” employed by C. Thi Nguyen and Mark Alfano, respectively. For examples of where the authors use these terms, see Alfano, “Virtues for agents,” 8427-28; Nguyen, “Epistemic Bubbles,” 150.

⁶⁷ This is an adapted version of Mark Alfano’s example of using one’s smart phone to learn about the 2015 Paris terrorist attacks from a testimonial network of news websites. Alfano, “Virtues for agents,” 8427-28.

members.⁶⁸ Trust-manipulation by Karl's epistemic network will be an alternative explanation of what happened to him. Finally, in section 4.3, I will argue this alternative explanation pose two epistemic structuralist challenges for HIVR.

4.1: A psychological model of testimony assessment online: background beliefs and trust

As we shall see in section 4.2, background beliefs and trust play a major part in Nguyen's account of how agents like Karl are led astray by trust manipulation by echo chambers.⁶⁹ Therefore, to prepare for this account, this section provides a psychological model emphasizing the role of background beliefs and trust in assessing expert testimony online. The model is primarily inspired by my interpretation of Nguyen's work.⁷⁰ First, I will argue why the Web accessible via search engines are better conceptualized as a testimonial space than an information space. Then, I will present the psychological model whose two components—background beliefs and trust—are used to evaluate written expert testimony online.

The relevance of the model becomes clear once we reconceptualize the Web as a “testimonial space.” Recall HIVR conceptualized the Web as an information space. Yet, I contend the case of Karl shows this conceptualization is not the most accurate one. We saw Karl did not only considered the information contained in a statement about the effectiveness of the COVID-19 vaccine. More importantly, he considered *who* made the statement. For this reason, I maintain it is more accurate to say Karl was assessing, not just “mere” information, but written expert testimony about COVID-19.⁷¹ By ‘written expert testimony’ (‘testimony’, for short), I mean a statement written by an expert providing evidence that *p* is true to a reader who wants to know.⁷² It follows from the definition testimony has two components: content and source.⁷³ For example, when Karl encountered the statement “the COVID-19 vaccine is dangerous,” this was the content and the natural medicine expert the source. Webpages Karl encountered are more accurately conceptualized as a collection of testimony about COVID-19 vaccines.⁷⁴

⁶⁸ Nguyen, "Epistemic Bubbles."

⁶⁹ Nguyen, "Epistemic Bubbles."

⁷⁰ Nguyen, "Epistemic Bubbles."; Nguyen, "Cognitive islands."

⁷¹ Other authors have also noted the testimonial nature of the Web and social media. See for example Alfano, "Virtues for agents."; Rini, "Fake News."

⁷² I am echoing Miranda Fricker's formulation of Coady's definition of testimony. Cf. Fricker, "What's the Point of Blame?," 172. As Leonard's SEP entry brings out, there is disagreement in the literature about the nature of testimony. Leonard, "Epistemological Problems of Testimony." The entry does not discuss written testimony, I find it reasonable that there is written testimony. Like its spoken counterpart, a written statement has content and a source. This chapter assumes written testimony works similarly to its spoken counterpart.

⁷³ Here I follow Sperber et al.'s distinction between the source and content of testimony. They distinguish between: "[...] vigilance towards the source of communicated information and vigilance towards its content." Sperber et al., "Epistemic Vigilance," 367.

⁷⁴ My testimonial understanding of websites makes two simplifying assumptions. First, I assume the agent only encounters written testimony. Second, I assume web pages only have one author (e.g., they are published on

Heersmink's account acknowledges that a virtuous agent assesses sources, but does not couch his point in terms of testimony.⁷⁵ For the reasons given above, I maintain it is more accurate to conceptualize the Web as a testimonial space than an information space.

With a reconceptualization of the Web as a testimonial space, let us consider the two components of the psychological model of testimony assessment. The first component of the model is background beliefs. As I interpret Nguyen, agents assess testimony using background beliefs.⁷⁶ I want to distinguish between two types of background beliefs involved in testimony assessment: content beliefs and source beliefs.⁷⁷ This distinction corresponds to the two components of testimony we made above, namely content and source. Content beliefs are background beliefs an agent employs to assess the *content* of testimony. An example would be the background beliefs about natural medicine Karl invokes to assess the content of testimony about the COVID-19 vaccine. Source beliefs are background beliefs invoked in assessment of the *source* of testimony. Among these are background beliefs about what makes a testifier a good source of information in general (e.g., reliability). I also place what Nguyen calls "credentialing beliefs" among source beliefs.⁷⁸ He does not provide a formal definition of "credentialing belief." Judging from his use, however, I take it mean a belief about what makes a person an expert in a knowledge domain.⁷⁹ For example, a credentialing belief about the knowledge domain of medicine could be "the source has extensive experience and training in natural medicine."

The second component of the model of testimonial assessment is trust. This component is related to the first. As I interpret Nguyen, background beliefs—especially credentialing

their personal website). Non-written testimony and multi-author web pages introduce technicalities whose treatment is outside the scope of this chapter.

⁷⁵ Evidence of this is that Heersmink's discussion of how a virtuous agent uses search engines makes no mention of assessing testimony. Cf. Heersmink, "Virtue Epistemology," 6-8.

⁷⁶ Several accounts emphasize the role of (background) beliefs in assessment. In one paper, Nguyen writes that background beliefs are used to "vet new sources" and to assess evidence. Nguyen, "Epistemic Bubbles," 156-57. In another paper, Nguyen's discussion of novices also emphasizes the role of "starting beliefs" in assessing experts. Nguyen, "Cognitive islands," 2819. For another account emphasizing the role of background beliefs in testimonial assessment (primarily assessment of content), see Sperber et al., "Epistemic Vigilance," 374-76.

⁷⁷ The distinction between content beliefs and source beliefs are inspired by Sperber et al.'s distinction between assessing the content and source of testimony (see footnote 73 above). For a discussion of vigilance towards the source of testimony, see Sperber et al., "Epistemic Vigilance," 369-71. For a discussion of vigilance towards the content of testimony, see Sperber et al., "Epistemic Vigilance," 374-76.

⁷⁸ Nguyen, "Epistemic Bubbles," 157.

⁷⁹ In addition to "credentialing belief," Nguyen also introduces the notion of a "credentialing structure" in his discussion of echo chambers. Nguyen, "Epistemic Bubbles," 157. However, it is unclear to me whether "credentialing structure" refers to a structure in the *agent's* credentialing beliefs or whether it is a structure in the *social structure* echo chamber. In the absence of a clear definition of 'credentialing structure, I will stick to credentialing beliefs in this chapter.

beliefs—are decisive for trust allocation.⁸⁰ The emphasis upon credentialing beliefs is reasonable. An agent considers an expert in a domain trustworthy, only if the expert aligns with the agent’s beliefs about what constitutes an expert in that domain. Credentialing beliefs are especially important for novices like Karl, who form many of his beliefs about COVID-19 on testimony. A novice like Karl has few content beliefs about vaccines, unlike a vaccine expert. Thus, if Karl forms beliefs from the content of a piece of testimony about the COVID-19 vaccine, this is largely because the testifier satisfies his source beliefs—especially his credentialing beliefs about vaccines.⁸¹ By contrast, for a vaccine expert to believe on testimony, the testifier must satisfy his credentialing beliefs, *as well as* cohere with his content beliefs about vaccines. To be clear, novices also employ content beliefs in testimonial assessment. But they have fewer content beliefs than the expert to invoke in testimonial assessment.

In sum, in this section we have seen the Web is most accurately considered a testimonial space. We have also seen a psychological model of assessment whose two components—background beliefs and trust—are used to assess written expert testimony.

4.2: Alternative explanation: the agent was led astray by rigged background beliefs

Having sketched a psychological model of testimonial assessment, we are better prepared to understand how Nguyen’s analyses can explain what happened to Karl. I will argue the agent was led astray not by being epistemically vicious. Rather, I will argue Karl was led astray because he was situated in an echo chamber manipulating his trust by rigging his background beliefs in favor of the echo chamber. I will use Nguyen’s analyses to construct an argument against HIVR’s claim about epistemic character being decisive for good information-seeking

⁸⁰ I see evidence for this interpretation in several papers by Nguyen. One piece of evidence is the following quote: “People with different moral beliefs, commitments, and sensibilities will pick different experts in whom to put their greatest trust.” Nguyen, “Cognitive islands,” 2816. Similarly, I see evidence for this interpretation in Nguyen’s remark about escaping the trust-manipulation of echo chambers: “In order to undo the influence of historical ordering, an epistemic agent will have to temporarily suspend belief in all their beliefs, *in particular their credentialing beliefs* [Nguyen’s italics], and start from scratch.” Nguyen, “Epistemic Bubbles,” 157. As I understand him, Nguyen here emphasize credentialing beliefs because they play a pivotal role in trust allocation. Smart and Clowes also brings in trust-related beliefs in their discussion of virtue responsibility, extended knowledge and Google Search. Smart and Clowes, “Intellectual Virtues,” 18.

⁸¹ Nguyen makes a similar point with his concept of “runaway bootstrapping”: “[...] the experts I put my trust in will be a reflection of, and an extension of, my own abilities, commitments, and sensibilities in that domain.” Nguyen, “Cognitive islands,” 2817. Note that Nguyen’s claim only applies to knowledge domains where there are no empirical expertise test. A similar phenomenon (“runaway credence levels”) happens when the agent’s trust is manipulated by an echo chamber. Nguyen, “Epistemic Bubbles,” 150.

and -assessment behavior when using search engines.⁸² I will do so by first arguing Karl's epistemic network satisfies Nguyen's definition of an echo chamber, before proceeding to explain how trust-manipulation of a natural medicine echo chambers can explain why Karl was led astray.⁸³

First, let us show that the agent's epistemic network satisfies Nguyen's definition of an echo chamber. Nguyen's describes echo chambers generally as "[...] an epistemic community which creates a significant disparity in trust between members and non-members."⁸⁴ Nguyen specifies his definition further by highlighting two components: membership beliefs and what I will call "trust-manipulation mechanisms".⁸⁵ Although it might not be obvious that a network of web pages can be an echo chamber, I will argue that the network nonetheless satisfies the two components of the definition.

The first component of echo chambers is agreement with a set of beliefs, some of which support distrust in non-members.⁸⁶ In my terms, I take this to mean echo chambers require agreement with a set of content and source beliefs conducive to distrust. A natural medicine echo chamber, for instance, likely requires members to agree with the content belief that natural medicine is superior to traditional medicine. Plausibly, the epistemic network of Karl required this. Also, it is likely a natural medicine echo chamber requires members to agree to source beliefs about the untrustworthiness of people hostile to natural medicine. For example, as we saw in the case above, one natural medicine expert in Karl's epistemic network claimed vaccine advocates are part of a depopulation plot. Thus, in this way, it is likely Karl came to accept and agree with beliefs making him a member of an echo chamber.

Karl's epistemic network also satisfies the second component of the echo chamber definition: use of trust-manipulation mechanisms. I have chosen to call these "trust-manipulation mechanisms" because I see them as crucial to Nguyen's account of the trust-manipulation echo chambers do. Nguyen identifies three trust-manipulation mechanisms: Epistemic discrediting of non-members, amplifying epistemic credentials of members, and disagreement-reinforcement mechanisms.⁸⁷ Epistemic discrediting is a mechanism whereby

⁸² See Nguyen's own engagement with Quassim Cassam's virtue epistemological account. Nguyen, "Epistemic Bubbles," 155-56. Although he seems to consider epistemic vices relevant to echo chamber cases, he is skeptical to *individualist* virtue epistemological accounts of echo chamber cases.

⁸³ This part of my argument draws upon the analyses of echo chambers in Nguyen, "Epistemic Bubbles."

⁸⁴ Nguyen, "Epistemic Bubbles," 146.; Nguyen's italics removed.

⁸⁵ Nguyen, "Epistemic Bubbles," 146-47.

⁸⁶ Nguyen, "Epistemic Bubbles," 146.

⁸⁷ Nguyen, "Epistemic Bubbles," 146-47. On Nguyen's view, the disagreement-reinforcement mechanism is not considered essential to echo chambers. My argument does not hinge on the mechanism being essential or not.

non-members “[...] are actively assigned some epistemic demerit, such as unreliability, epistemic maliciousness, or dishonesty.”⁸⁸ We saw Karl’s epistemic network assigned dishonesty to vaccine advocates because they are part of a depopulation plot. Amplification of epistemic credentials is a mechanism whereby “[...] members are assigned very high levels of trust.”⁸⁹ We saw one expert in Karl’s epistemic network claimed only *they* know the “real truth” about COVID-19 vaccines. This implication is only people who know the “real truth” are trustworthy sources of information about vaccines. Finally, disagreement-reinforcement mechanism instills agents with “[...] a set of beliefs such that the existence and expression of contrary beliefs reinforces the original set of beliefs and the discrediting story.”⁹⁰ Recall the expert in Karl’s epistemic network claiming vaccine advocates will deny the dangers associated with the COVID-19 vaccine. And we saw Karl considers this verification of the expert’s claim about the untrustworthiness of vaccine advocates like WHO. Because Karl’s epistemic network employed all three trust-manipulation mechanisms, it also satisfies the second component. Hence, since both components are satisfied, the agent’s epistemic network is an echo chamber.

Knowing the epistemic network is an echo chamber, Nguyen’s analysis, sketched above, of the influence of echo chambers upon agents can provide an alternative explanation of what happened to Karl. Through trust-manipulation by echo chambers, Nguyen notes, “[...] members are not just cut off, but are actively alienated from any of the usual sources of contrary argument, consideration, or evidence.”⁹¹ To my understanding, this happens because the echo chamber rigs the members’ background beliefs in favor of the echo chamber. Nguyen appears to think rigging of credentialing beliefs as particularly decisive for trust-manipulation by echo chambers.⁹² This is reasonable. Recall the psychological model: Whether an agent trusts someone as an expert on a topic will depend on whether they satisfy the agent’s credentialing beliefs. And if credentialing beliefs are rigged in favor of an echo

⁸⁸ Nguyen, "Epistemic Bubbles," 146.

⁸⁹ Nguyen, "Epistemic Bubbles," 146.

⁹⁰ Nguyen, "Epistemic Bubbles," 147.

⁹¹ Nguyen, "Epistemic Bubbles," 147. It is unclear to me whether Nguyen claims agents situated in echo chambers *truly* evaluate the claims they disagree with. On the one hand, as we see in the quote referenced, Nguyen writes echo chamber members are “actively alienated” from contrary views and evidence because their trust is manipulated by the echo chamber. However, in a later discussion of an agent raised in an echo chamber, acting virtuously in many ways, Nguyen writes that they use their background beliefs to assess sources and evidence. Cf. Nguyen, "Epistemic Bubbles," 155, 157. I will assume unvirtuous echo chamber members might not assess contrary views, but more virtuous echo chamber members will. Karl, I have argued, belongs to this category of virtuous echo chamber members.

⁹² Revising credentialing beliefs is highlighted in Nguyen’s solution for undoing the influence of echo chambers (see footnote 80 above).

chamber, the agent will only trust as expert sources compatible with and endorsed by the echo chamber.

Crucially, Nguyen claims that when an agent's is situated in an echo chamber, the echo chamber's influence makes the agent's "epistemic practices" unreliable.⁹³ The vice is not part of individual members such as Karl, but of the echo chamber.⁹⁴ To my understanding, this is because the *echo chamber* manipulates and maintains the rigged background beliefs, guaranteeing Karl will not assess contrary views as untrustworthy. And as a result, Karl will not assess them as more accurate or well-founded than their rigged background beliefs. This provides us with an alternative explanation of what happened to Karl. He was led astray not because they were epistemically vicious, such as lacking open-mindedness. Rather, the agent was led astray because he was situated in a natural medicine echo chamber, manipulating his trust via rigging of background beliefs. His rigged background beliefs made him find and assess testimony on web pages in a way that guarantees vindication of the natural medicine echo chamber. Through influence of echo chambers, an agent can be virtuous in HIVR's sense, yet still engage in bad search engine behavior.

4.3: Two epistemic structuralist challenges for Heersmink's virtue responsibilism

The alternative explanation constitutes an epistemic structuralist argument, posing two challenges for HIVR. It is an epistemically structuralist argument because it puts explanatory and practical emphasis upon echo chambers, a social epistemic structure. Karl's search engine behavior was explained as largely the result of the trust-manipulating activity of the natural medicine echo chamber. The vice was identified with the echo chamber, not Karl. This suggests improvements in search engine behavior requires changes in this social epistemic structure. For instance, by adding a trustworthy non-member to the echo chamber.⁹⁵

⁹³ Nguyen, "Epistemic Bubbles," 155.

⁹⁴ Nguyen, "Epistemic Bubbles," 155. Two remarks about the passage referenced here are in place. First, Nguyen frames the point of vice being a part of the "collective intelligence". But I think the point still stands without invoking the concept of collective intelligence. To avoid complications, I will not introduce it here. Second, it is not apparent to me whether Nguyen would claim echo chamber members (especially those raised in them) are *fully* virtuous or only having *many* virtues. On the one hand, in the scenario with a person raised in an echo chamber, Nguyen describes him as seemingly "[...] behaving with many epistemic virtues." Nguyen, "Epistemic Bubbles," 155. On the other hand, his claim about echo chambers "[...] convert[ing] individually epistemically virtuous activity into collective epistemic vice" also admits of a stronger reading. Nguyen, "Epistemic Bubbles," 155. The agent can be fully virtuous while situated in an echo chamber. As I argue Karl satisfies the virtues listed by Heersmink's virtue responsibilist account discussed above, I assume the stronger reading.

⁹⁵ Cultivating a trust relationship to a non-member is crucial to escaping an echo chamber, according to Nguyen. Nguyen, "Epistemic Bubbles," 158. See also Mark Alfano's discussion of various ways topological improvements can be made to improve testimonial networks. Alfano, "Virtues for agents," 8432-36.

Thus, the epistemic structuralist argument presents two challenges for HIVR. First, it presents an *explanatory challenge* (EC): How can HIVR explain the epistemic outcome of the virtuous agent Karl in the natural medicine echo chamber, without emphasizing background beliefs over epistemic character? Recall virtue epistemology emphasize cognitive properties, such as epistemic character, more than beliefs. Since Karl is virtuous in HIVR's sense, we would expect him to have a good epistemic outcome. But we saw the virtuous agent Karl reliably failed to do so. Second, the argument presents a *practical challenge* (PC): If Karl is already virtuous, how (if at all) can the agent improve his search engine behavior when he is situated in the natural medicine echo chamber? Karl is already virtuous in HIVR's sense. Improving his epistemic character is seemingly not an option to improve his behavior.⁹⁶

In the next section, I argue an individualist virtue responsibilism can *in principle* respond to these challenges by adding a trust-related virtue.

5: A Trustworthy Virtue Responsibilism for the Internet?

The crux of the epistemic structuralist argument is Karl directs his trust badly because his trust was manipulated by an echo chamber. HIVR runs into problem because it underestimates the importance of trust when using search engines, more specifically finding good experts to trust.⁹⁷ The problem, then, seems to be HIVR's list of virtues lacks a trust-related virtue. By adding a trust-related virtue, HIVR should in principle be able to meet both the explanatory and practical challenge.⁹⁸ I call this version of virtue responsibilism the "augmented individualist virtue responsibilism" ("AIVR," for short). Adding another virtue is compatible with HIVR, which admits its list of virtues is not exhaustive.⁹⁹ In section 5.1, I provide brief sketch of a trust-related virtue I call "the virtue of well-directed trust" ("WDT,"

⁹⁶ This challenge is inspired by a similar challenge in an argument in a paper by Michel Croce and Tommaso Piazza. Croce and Piazza, "Consuming Fake News." One of the premises is that users of social media do not act in an epistemically blameworthy fashion when consuming fake news on social media. Croce and Piazza, "Consuming Fake News," 2. C Thi Nguyen's notion of "runaway bootstrapping," where flawed beliefs can lead an otherwise impeccable agent astray, also have similarities to the challenge (see footnote 81).

⁹⁷ On the challenge of deferring to reliable authority online, Heersmink writes: "[...] being intellectually autonomous does not mean that one cannot ask others for advice. One may, for example, have read something on a medical webpage but still ask one's doctor whether the information is correct and the source reliable." Heersmink, "Virtue Epistemology," 11n5. But here Heersmink assumes the agent already has a true belief about which expert to ask for advice. The case of Karl has shown this to be a problematic assumption.

Also, I am not the first to highlight the importance of trust in an online setting. See for example Alfano and Klein, "Trust.," O'Neill, "Trust." See especially Mark Alfano's work engaging with "virtuous (dis)trust" in directed social networks. Alfano, "Virtues for agents," 8431. I return to this account in footnote 100.

⁹⁸ Croce and Piazza also aim to defend "educational" approaches against the "structural approaches" to fake news consumption. Croce and Piazza, "Consuming Fake News," 8. Unlike my account, however, Croce and Piazza explicitly restrict their educational solution applies to "ordinary social media users," not agents situated in echo chambers. Croce and Piazza, "Consuming Fake News," 8.

⁹⁹ Heersmink, "Virtue Epistemology," 4.

for short). I argue it is a virtue responsibilist virtue. I also identify a crucial assumption for the response to work, namely that trusting well is most plausibly accounted for in individualist terms. Proceeding on this assumption, sections 5.2 and 5.3 argue the sketch of WDT makes AIVR meet the explanatory challenge (EC) and practical challenge (PC), respectively.

5.1: Sketch of the virtue of well-directed trust into individualist virtue

The individualist response to the epistemic structuralist argument I gave above was to add WDT. By adding WDT, the idea goes, an agent will gather information online well by directing their trust well, such as towards good experts. The idea is that if Karl had the character trait WDT, he would have trusted the WHO, not the natural medicine expert.

However, this responds raises important philosophical questions about trust in online information gathering. What *is* trust? What does it mean to trust well online? And most crucially for the individualist response to be effective: Can trusting well plausibly be accounted for in terms of the epistemic individuals (e.g., as a virtue like WDT)? These questions are beyond the scope of this chapter. I return to them in chapter 2. For the sake of this chapter, though, I assume trusting well can be accounted for in terms of WDT. The rest of this subsection provides a sketch of WDT in terms of the individual and argue it is a virtue responsibilist virtue.

For the sake of this chapter, I propose to understand WDT as, roughly, a disposition of an agent to trust someone (e.g., an expert) with respect to a knowledge domain to the extent to which they are trustworthy (e.g., reliable, knowledgeable, honest) in that domain.¹⁰⁰ The virtue has two aspects. First, it is a disposition to deliberate well with respect to the trustworthiness. For example, an agent who possesses WDT reasons well about how to find

¹⁰⁰ The sketch is the result of my engagement with numerous sources. I am influenced by the feminist virtue epistemology literature, who have proposed trust-related virtues. See for example Daukas, "Epistemic Trust."; Daukas, "Altogether Now." An important thinker for this sketch has been Miranda Fricker's formulation of testimonial justice as an Aristotelean virtue to give testimonial sources the credence they deserve. Fricker, *Epistemic Injustice: Power and the Ethics of knowing*, 92-93. My sketch of WDT differs in important respects. It is less focused on the harm done by epistemic injustice. Also, my account does not emphasize prejudice as much as Fricker does. Although directing trust based on prejudice can lead to badly directed trust and harm, my focus is on the epistemic loss experienced by the director of trust. I flesh this out in chapter 2, where I define WDT as a virtue responsibilist virtue.

I also want to differentiate my account of WDT from Mark Alfano's account of virtues in directed social networks. Alfano, "Virtues for agents." As mentioned in footnote 96, he briefly mentions virtuous trust in an online context. He also connects his account to Fricker's notion of testimonial justice. Alfano, "Virtues for agents," 8437, 8439. While Alfano's account is certainly relevant, it seems to me to be an epistemic structuralist account of trust. However, here I am interested in formulating a strictly epistemic individualist account of a trust-related virtue. Only an epistemic individualist account can save HIVR from the epistemic structuralist critique. Therefore, I will not pursue Alfano's account of virtues in directed social networks further here and in chapter 2.

and assess the trustworthiness of experts using search engines.¹⁰¹ The second aspect is being disposed to make good intuitive judgements about trustworthiness. This is meant to capture an Aristotelean idea that the virtuous agent will immediately “perceive” the correct thing to do in a situation, such as whether someone should be trusted as a source of information.¹⁰² One way to think about this is a “hunch” or “gut feeling” that someone is trustworthy. Sometimes, the intuitive judgement is refined and made explicit through deliberations. Other times, it might be hard to put the intuitive judgement into words. A novice with WDT might not be able to explain why the natural medicine expert is not a trustworthy expert with respect to vaccines, but will nonetheless feel something is “off” about their trustworthiness.

This sketch of WDT plausibly qualify as a virtue responsibilist virtue in the vein of HIVR. Based on the presentation of virtue responsibilism and HIVR, I suggest a virtue responsibilist virtue should meet three criteria: it should be *acquirable*, *truth-conducive*, and *error-minimizing*. First off, it is plausible WDT can be acquired through education or practice. For example, in a classroom setting, an agent can practice their trust direction by using search engines to answer a question like “Did the moon landing in 1969 happen?”.¹⁰³ Discussion and reflection upon the sources they chose as trustworthy can help them cultivate a disposition to trust well. Second, it is plausible WDT is both truth-conducive and error-minimizing. Forming beliefs by trusting reliable experts will, on average, make the agent acquire more true beliefs than false beliefs. Thus, there are good reasons to think the sketch of WDT is a virtue responsibilist virtue.

In the next two subsections, I will argue including WDT helps AIVR address the explanatory challenge (EC) and the practical challenge (PC).

5.2: Addressing the explanatory challenge (EC)

Recall (EC) was the challenge of how HIVR can explain the epistemic outcome of the agent situated in a natural medicine echo chamber, without emphasizing background beliefs over epistemic character.

AIVR *can* address (EC). AIVR can argue the bad background beliefs are an *effect* rather than the *cause* of badly aimed trust. The cause of badly aimed trust is having an epistemic

¹⁰¹ For an account of how laypersons can use an online resource like Wikipedia to assess trustworthiness of experts, see Anderson, "Democracy," 149-53. For a discussion of how novices can assess experts by assessing their track-record of experts, see Goldman, "Experts," 106-08.

¹⁰² Thanks to Sebastian Watzl for suggesting the intuitive part for the Aristotelean parallel.

¹⁰³ The relevance of this question and inspiration for the practical exercise I have from an article by Gloria Origgi. Origgi, "Epistemic Injustice and Epistemic Trust," 222-23.

character disposed to trust unreliable experts.¹⁰⁴ AIVR can argue Karl was *not* fully virtuous. He lacked a crucial epistemic virtue: WDT. Lacking this virtue, he did not possess an epistemic character disposed to trust reliable experts. Had he possessed the virtue, his epistemic character would have been disposed to consider WHO a more trustworthy source than the natural medicine expert. First, he would likely have an immediate judgement that the WHO is an actually reliable source about the COVID-19 vaccine. Second, his deliberative assessment will likely recognize hallmarks of genuine expertise in WHO. For opposite reasons, he would assess the natural medicine experts untrustworthy, because they are not reliable experts on vaccines. Adding WDT, AIVR can provide an explanation of the epistemic outcome of Karl, without emphasizing background beliefs over epistemic character.

5.3: Addressing the practical challenge (PC)

Recall (PC) was the challenge of how Karl, virtuous in HIVR's sense, can improve his search engine behavior when he is situated in an echo chamber. I will argue that AIVR *can* address (PC) indirectly.

Consider this naïve response to (PC). Above we explained Karl's failure as his lacking WDT. Therefore, the naïve response goes, Karl should just cultivate WDT. With WDT, Karl will start to trust better sources and come to hold a more accurate view of the COVID-19 vaccine. He should then be disposed to recognize the WHO as a reliable expert about the COVID-19 vaccine, and the natural medicine expert a bad one. As a result, it is plausible the beliefs he acquires via trusting the WHO will be more accurate. In this way, the naïve response contends Karl will improve his search engine behavior simply by cultivating WDT. Insofar as epistemic practicability implies responsibility, Karl will also be responsible for doing so.

The problem with the naïve response is that it assumes it is epistemically practicable for agents to cultivate WDT in echo chambers. While the naïve response works in principle, there are good reasons to consider this assumption implausible in practice. For one, it is highly implausible the agent will realize their beliefs are mistaken; echo chamber members believe they have good explanations for the testimony they encounter.¹⁰⁵ They believe they know the "real truth" about COVID-19 vaccines. For another, as we saw, echo chamber members are likely interpreting resistance to their echo chamber beliefs as vindicating them. If we told Karl he trusts bad experts, he would likely maintain that *he* trusts the good experts. Moreover, he

¹⁰⁴ According to Alfano and Klein, authors in epistemology have proposed vices related to distrust. Alfano and Klein, "Trust," 3. One could argue the agent situated in the echo chamber possesses a vice related to distrust.

¹⁰⁵ Nguyen, "Epistemic Bubbles," 158.

would likely consider our confrontation as evidence of us being brainwashed by the depopulation plot. It is unlikely Karl will realize he trusts badly, let alone recognize the need to cultivate WDT. Thus, it seems epistemically impracticable for him to cultivate WDT. For these reasons, the naïve response to (PC) fails.

Yet, AIVR might address (PC) more indirectly. It is likely epistemically impracticable for an agent to improve their search engine behavior *while situated* in an echo chamber. However, it might be epistemically practicable for the agent to cultivate their character *before* getting trapped in an echo chamber. Nguyen's points about agents raised in echo chambers are instructive for our discussion. He notes that agents who are *not* raised in echo chambers can be blameworthy for acquiring echo chamber beliefs; by contrast, agent who *are* raised in echo chambers are not blameworthy for their beliefs.¹⁰⁶ For example, if Karl was raised in a natural medicine echo chamber, he would plausibly not be blameworthy for holding his beliefs. His conclusion is: Provided evidence is considered using the problematic background beliefs, "[...] early education in an echo chamber becomes domineeringly powerful."¹⁰⁷

AIVR can make a similar point about education and epistemic character. This can provide a way for AIVR to meet (PC) preemptively. Karl was unlucky insofar as he was raised in a natural medicine community where the cultivation of WDT was epistemically impracticable. Karl engages in bad search engines behavior because his community did not provide him with the opportunity to cultivate WDT.

But the converse of this result provides an opportunity to meet (PC) indirectly. Had Karl been raised in a community where cultivation of WDT *was* epistemically practicable, he *could have* learned to trust well. If so, Karl would likely have judged the natural medicine experts as untrustworthy experts about the COVID-19 vaccine. By not finding echo chamber experts trustworthy, Karl could drastically reduce the probability of falling into an echo chamber. In this way, AIVR can address (PC) preemptively.

An epistemic structuralist could object that the preemptive solution smuggles in epistemic structuralism. They claim the decisive factor is the social structure in which Karl is raised. Without the educational opportunities to cultivate WDT, afforded to Karl by the formative social structure, Karl would not acquire the epistemic character to engage well with search

¹⁰⁶ Nguyen, "Epistemic Bubbles," 155-56.

¹⁰⁷ Nguyen, "Epistemic Bubbles," 157.

engines. Hence, the defender concludes, the practical emphasis ought to be on the formative social structure, not on epistemic the individuals raised in them.

AIVR can respond to this objection. AIVR can admit the formative social structure provide necessary educational *background conditions* for cultivation of epistemic character.¹⁰⁸ Yet, it does not necessarily follow the formative social structure ought to have primary practical emphasis. If we want to improve search engine behavior, AIVR could argue, we need to ensure agents like Karl live up to their responsibility to cultivate a good epistemic character. Background conditions in and of themselves will not produce good search engine behavior. Thus, AIVR could conclude, it is still reasonable to emphasize epistemic individuals with respect to improving search engine behavior. Thus, I conclude the epistemic structuralist failed to show the indirect response to (PC) implies epistemic structuralism.

Conclusion

This chapter tackled the question: Are good character traits sufficient for making Internet users solve the epistemically irrelevant information problem? We saw HIVR's answer was "yes," but that the challenges of the epistemic structuralist argument called this affirmative answer into question. However, I argued that AIVR, by adding WDT, can *in principle* address the challenges raised both the explanatory challenge and in principle the practical challenge. Yet, I noted that this response proceeds on a condition, namely the assumption that directing trust well is most plausibly accounted for in terms of the epistemic individual, such as a virtue responsibilist virtue. In the next chapter, I will discuss whether this condition holds.

¹⁰⁸ Nguyen also uses the notion of background conditions in his "conceptually minimal" claim about the epistemic influence of echo chambers on epistemic individuals. Nguyen, "Epistemic Bubbles," 156.

Chapter 2: The virtue of well-directed trust: an individualist or extended virtue?

Chapter 1 provided an epistemic individualist response on behalf of AIVR to the epistemic structuralist critique. The response's success rests on whether the following condition holds: well-directed trust is most plausibly accounted for primarily by individual factors, more specifically as an individualist virtue. This chapter will argue it does not hold.

As chapter 1's sketch showed, it is *prima facie* plausible to think of well-directed trust in the online environment in individualist terms. As the sketch of WTD showed us, trusting well might only need to appeal to individual factors like the agent's dispositions. However, as I will argue in this chapter, there are good reasons to think well-directed trust in the online environment involves more than individual factors. C. Thi Nguyen has argued trust is an important strategy for agents to manage their cognitive and practical limitations.¹ Based on this observation, it is often crucial for agents to trust online epistemic tools like search engines. An agent only has the resources to assess a fraction of purported experts on the Internet. By ordering information according to epistemic relevance, I argue, online epistemic tools play crucial role in trusting well online. But the importance of online epistemic tools suggests there are more than individual factors involved in trusting well. If so, this suggests the virtue of well-directed trust is most plausibly conceptualized as an *extended* virtue. I define 'extended virtue' as a virtue where at least one considerable cause of virtuous activity's result is located outside the individual. In this case, a cause is "considerable" when it exerts more causal influence than a mere background condition on trusting well.

The question I want to address in this chapter is: Is the virtue of well-directed trust most plausibly conceptualized as an individualist virtue or an extended virtue? This is an important question to answer for the epistemic individualist. There might exist a virtue for well-directed trust. Yet, if this virtue is best conceptualized as an extended virtue, then AIVR cannot appeal to this virtue to defend epistemic individualism against the epistemic structuralist critique.

There are several things I do not have the space to cover in this chapter. For one, I focus mainly on trust, not on related phenomena like distrust.² Only briefly, in section 3.2.2, will I consider distrust, which Nguyen understands as an "actively questioning attitude."³ For another, although parts of my discussion will revolve around causality, I do not have the space

¹ Nguyen, "Trust," 19, 45. See also footnote 16 of this chapter.

² For a brief overview of distrust, see McLeod, "Trust."

³ Nguyen, "Trust," 25.

to provide an elaborate discussion of theories of causality. Finally, my focus will be restricted to epistemic trust, not moral trust.⁴

Let me indicate the methodology and terminology I will use to explore this chapter's question. I will use the fictional case of Taylor. Taylor, like the agents in Chapter 1, has an epistemic goal of answering the question: How does the COVID-19 vaccine work? This question informs her decision about whether to take the vaccine. Also like the agents in Chapter 1, she is confronted with the problem of epistemically irrelevant information online. Like the others, she solves this problem by finding a good expert to trust using search engines. She has two tasks. First, she must find expert testimony about the COVID-19 vaccine. Her second task is to assess the trustworthiness of purported experts. But Taylor differs from the agents in Chapter 1. In addition to the virtues discussed in Chapter 1, she also has the disposition associated with the individualist conception of well-directed trust. The idea is to assess the plausibility of the individualist conception by checking whether this disposition is enough to make Taylor direct her trust well using search engines.

In this chapter, I employ a simplified understanding of a search engine. I make the simplifying assumption that the search engine orders web pages according to relevance using a single ranking algorithm. The ranking algorithm I assume to be a machine learning algorithm. Recall from the introductory chapter that this is roughly an algorithm using data to learn to make more accurate predictions—in this case about the epistemic relevance of web pages. Ideally, the ranking algorithm provides a ranking of web pages according to epistemic relevance. In this ideal situation, all Taylor needs to do to find a good expert is to assess the best expert on the first couple of pages.

However, in practice, search engines might be biased. In this chapter, I focus on a particular type of bias I call "bias towards epistemic irrelevance." Although search engines can be biased numerous ways, bias towards epistemic irrelevance is highly pertinent to the problem of epistemically irrelevant information online.⁵ Henceforth, unless stated otherwise, 'bias' refers to bias towards epistemic irrelevance. Insofar as bias towards epistemic irrelevance can place good experts lower in the web page ranking, bias towards epistemic irrelevance can be decisive in whether Taylor finding and trusting the best expert about the COVID-19 vaccine. While this chapter focuses mainly on search engines, many of the same challenges apply to

⁴ For a good overview of different notions of trust, see Origgi, "Is Trust an Epistemological Notion?"

⁵ The literature has focused on the implications of machine learning algorithms' bias towards social groups. As examples, see Gebu, "Race and Gender."; Johnson, "Algorithmic bias."

other online epistemic tools. As we will see in Chapter 3, the problem of bias towards epistemic irrelevance is also a problem for the Facebook News Feed.

This is the structure of the chapter. In section 1, I will provide what a plausible individualist conceptualization of the virtue of well-directed trust should look like. To this end, I use C. Thi Nguyen's unquestioning attitude account to explore what trust means in the context of online information gathering. I will also introduce what I call the capacity of directing trust. I will argue the individualist conception of the virtue of well-directed should be a disposition of the capacity to direct trust in a truth-conducive, error-minimizing manner. In section 2, I present my main argument that the virtue of well-directed trust is most plausibly conceptualized as an extended virtue. The argument makes two central claims. First, possessing the individual disposition identified will not necessarily make Taylor direct her trust well. Second, the search engine's bias exerts a considerable causal influence upon whom Taylor trusts. This, I argue, makes the extended virtue conceptualization the most plausible one. Section 3 raises an objection that my argument does not follow, because Taylor was in fact epistemically gullible for trusting a potentially biased search engine. In section 4, I argue this objection fails. Even if Taylor was epistemically gullible, her trust in the search engine was rational. Second, the alternatives to trusting search engines are problematic in their own way and can even lend support to the plausibility of the extended virtue conception. My thesis is the virtue of well-directed trust is best conceptualized as an extended virtue.

1: Trust, Directing Trust, and the Epistemic Individualist Conception of the Virtue of Well-Directed Trust

1.1: Trust as unquestioning attitude and its role in information gathering online

We want to understand the role of trust in novices' gathering of epistemically relevant information online. Knowing what trust is and what it means to trust well will be important in later sections, where I discuss what an epistemic individualist virtue of well-directed trust should look like.

When we discuss trust and online information gathering online, we should appreciate the cognitive and practical limitations of agents. Agents are cognitively and practically limited beings.⁶ For example, as Jessie Munton has noted, human beings face cognitive limitations in

⁶ For the cognitive finitude of human beings, see Nguyen, "Trust," 2, 19, 32-33. In the finalization of this chapter, I discovered Nguyen elsewhere discusses cognitive limitations in the context of dealing with a barrage of relevant information. Nguyen, "Seductions of Clarity," 232-34, 233n4.

The claim about cognitive limitations of humans has support in many domains. The empirical researchers Chun, Golomb and Turk-Browne have argued attention is crucial because human beings have limited

their ability to acquire and process information.⁷ We would expect these limitations impact information gathering online. To indicate the plausibility of this claim, consider the case of Taylor. She is practically limited insofar as there is a finite number of hours per day she can devote to information gathering online. Spending many hours every day risks encroaching on other life obligations. Taylor also faces cognitive limitations in her information gathering online. Finding and assessing information is a cognitively demanding process. This puts restrictions on the effectiveness of their information gathering online.

According to C. Thi Nguyen, trust plays an important role in helping agents deal with their cognitive and practical limitations.⁸ Nguyen determines trust as an unquestioning attitude, involving two related dispositions.⁹ First, the unquestioning attitude implies the agent is disposed to “immediately accept” that the trustee performs a given function, such as providing information.¹⁰ If Taylor trusts the WHO as a source of epistemically relevant information about the COVID-19 vaccine, she will be disposed to accept their testimony.¹¹

Adopting an unquestioning attitude implies the agent also adopts a “second-order disposition”: the agent is disposed not to deliberate about having the first disposition, namely the disposition to accept that the trustee performs a given function.¹² If Taylor trusts the WHO, she will also be disposed to not deliberate about her accepting the WHO’s testimony about the COVID-19 vaccine. To my understanding, the “second-order disposition” makes her acceptance of WHO’s testimony robust.¹³ Even if the WHO were to make the occasional

capacities for information processing. Chun, Golomb, and Turk-Browne, "Taxonomy of External and Internal Attention," 75. In philosophy, Jessie Munton, citing empirical literature, has also emphasized cognitive limitations in information acquisition. Munton, "Perceptual Skill," 134. I hypothesize management of cognitive resources has been crucial for survival. For an epistemological account emphasizing survival, see Levy and Alfano, "Knowledge From Vice."

Also relevant here is Gabrielle M. Johnson’s discussion of biases as necessary for finite human beings to reduce the space of possible theories about the world compatible with their evidence. Johnson, "Structure of Bias," 1195-96. In section 1.2 and footnote 45, I briefly return to the topic of trust-related biases and heuristics.

⁷ Munton, "Perceptual Skill," 134.

⁸ This is an interpretation built on various textual evidence. For trust as a “strategy” to manage cognitive limitations, see Nguyen, “Trust,” 2, 19. Elsewhere, Nguyen describes trust as a “response” to cognitive and practical limitations. Nguyen, “Trust,” 45. As I interpret Nguyen, trust is a strategy for dealing with *both* cognitive and practical limitations.

⁹ Nguyen, “Trust,” 19-20.

¹⁰ Nguyen, “Trust,” 19-20.

¹¹ I have adapted a similar example by Nguyen of trusting a news site. Cf. Nguyen, “Trust,” 40.

¹² Nguyen, “Trust,” 19-20.

¹³ Cf. Nguyen, “Trust,” 20. Here, Nguyen describes trust as “resistant”. This is what I take Nguyen to mean when he writes “inertia of trust.” Nguyen, “Trust,” 22.

mistake, this will not disrupt her disposition to accept the WHO's testimony about the COVID-19 vaccine.¹⁴

This is not to say that Taylor cannot lose her disposition not to question the WHO. An unquestioning attitude only implies that strong reasons are necessary to cease accepting the trustee's output.¹⁵ Should the WHO turn out to provide obviously epistemically irrelevant information on a consistent basis, Taylor would surely lose her disposition to accept the WHO's testimony.

How can trust, understood as an unquestioning attitude, help agents address their cognitive and practical limitations? As I interpret Nguyen, when an agent adopts an unquestioning attitude towards other agents and objects, they "outsource" their agency to agents and objects in the environment.¹⁶ Outsourcing practically and cognitively demanding tasks in this way, I take Nguyen to claim, liberates cognitive and practical resources the agent can spend on other tasks.¹⁷ Once Taylor has assessed the WHO as a trustworthy source of epistemically relevant information, accepting the WHO's testimony means she can spend her limited practical and cognitive resources on other tasks. Of course, this is only a good strategy, provided the WHO is a trustworthy source of epistemically relevant information. I will return to trustworthiness assessments in the next section. For now, the crucial point is this: Trust allows agents to manage their cognitive and practical limitations by outsourcing tasks to the environment.

Before summing up the role of trust in Taylor's online information gathering, we should note that the unquestioning attitude account rejects a common view that trust is directed only at agents.¹⁸ Nguyen wants to reject this view.¹⁹ Throughout the essay, Nguyen gives examples of how we adopt an unquestioning attitude towards various objects.²⁰ Insofar as we can outsource tasks to objects, objects should in principle also be targets of trust. But as Nguyen's

¹⁴ Cf. Nguyen, "Trust," 22. Nguyen writes: "The inertia of trust can survive the occasional disturbance." Nguyen, "Trust," 22.

¹⁵ Nguyen, "Trust," 20.

¹⁶ This interpretation is based on the following textual evidence. For the idea of trust as a strategy for managing an agent's cognitive and practical limitations, see Nguyen, "Trust," 2, 19. For the notion of trust as outsourcing agency, see Nguyen, "Trust," 42-43.

¹⁷ Nguyen's definition of trust *qua* unquestioning attitude is inspired by Richard Holton's work on intentions and resolutions. Nguyen, "Trust," 20. According to Nguyen, Holton claims that intentions and resolutions make agents "[...] settle their minds about some things in order to *free up cognitive resources for other projects* [my italics]." Nguyen, "Trust," 18. Seeing as Nguyen's unquestioning account on Holton's, I interpret Nguyen as claiming trust similarly liberates cognitive and practical resources.

¹⁸ Nguyen, "Trust," 5.

¹⁹ Nguyen, "Trust," 5. More precisely, he rejects that proper trust requires the trustee to be a bearer of "agential states."

²⁰ For an overview of examples of the variety of objects, see Nguyen, "Trust," 36.

claim is a controversial view, I should briefly outline the most relevant parts of Nguyen's argument for why trust in objects is legitimate.

Nguyen takes himself to be arguing against the idea that trust is only properly directed at agents.²¹ Nguyen adopts the same starting point as some agent-centered accounts, namely that the possibility of evoking a feeling of betrayal is the hallmark of trust.²² However, Nguyen reaches a different conclusion than agent-centered accounts. According to Nguyen, agent-centered accounts argue only *agents* can evoke a feeling of betrayal: only an agent can evoke a feeling of betrayal by, say, failing to express goodwill, be responsive to the truster, or fail to live up to commitments and promises made.²³ Since objects fail to perform these agential tasks, Nguyen takes agent-centered accounts to have concluded objects cannot evoke a feeling of betrayal: When objects fail us, they evoke, at most, a feeling of disappointment.²⁴ For these reasons, agent-centered accounts have concluded—wrongly, on Nguyen's view—that objects are not appropriate targets of trust.²⁵

Nguyen's unquestioning attitude account aims to show the agent-centered accounts are wrong about trusting objects.²⁶ According to Nguyen, trust and betrayal is intimately linked to "agential integration" and intimacy.²⁷ To my understanding, Nguyen claims betrayal can be a reaction when something we have made *intimately* integrated into our agency fails us.²⁸ For example, a band-member can be an intimate part of my agency *qua* musician. When a band-member breaks their promise about arriving to the concert on time, I will react with a feeling of betrayal. A crucial part of my agency *qua* musician, taken as a whole, has failed to perform its function.

The agent-centered accounts, I take Nguyen to claim, fail to appreciate that coordination with other agents—making commitments, say—is only one way to integrate someone into one's

²¹ Nguyen, "Trust," 5.

²² Compare to Nguyen's overview of agent-centered accounts of trust. Nguyen, "Trust," 3-5. Betrayal also plays an important role in his own account. Nguyen, "Trust," 5.

²³ Nguyen, "Trust," 4-5.

²⁴ Nguyen, "Trust," 3.

²⁵ Nguyen, "Trust," 5.

²⁶ Nguyen, "Trust," 5.

²⁷ Nguyen, "Trust," 28, 38.

²⁸ Nguyen, "Trust," 28. The difference between what we may call "moral trust" and the "integrative trust" is to me the least clear part of Nguyen's account. To my understanding, the difference is moral trust involves adopting a "participant stance" towards the trustee, whereas the "integrative trust" involves adopting an "integrative stance" towards the trustee. Nguyen, "Trust," 29. To my understanding, the difference between the stance we take towards agents understood as *participating* with us in moral attitudes and behaviors, and the stance we take towards agents and objects we try to *integrate* into our agency. Cf. Nguyen, "Trust," 29. As moral trust is not the focus of this chapter and thesis, I will not try to ascertain the difference between these different forms of trust.

agency.²⁹ One can also integrate by adopting an unquestioning attitude.³⁰ On this view, by not questioning that someone performs a particular function, they can become an integrated part of our agency. This form of integration requires no coordination with other agents.

Crucially, Nguyen argues we can adopt an unquestioning attitude towards *objects*, not just agents.³¹ To illustrate, suppose John has a pacemaker. John's agency crucially depends on unquestioningly accepting that the pacemaker will save him. If he questions the pacemaker, he will be riddled with fear, rendering him effectively incapacitated. Now, suppose the pacemaker failed, but John arrived at the hospital in time to be saved. Nguyen's account suggests John would react with betrayal toward the pacemaker for failing him.³² On Nguyen's view, this is because the pacemaker was intimately integrated into John's agency. The bottom line of Nguyen's overall argument, as I understand it, is this: Objects are appropriate targets of trust *qua* unquestioning attitude, because trusting objects help agents deal with their cognitive and practical limitations.

With an understanding of why objects—like search engines—are appropriate targets of trust, let us consider the upshot of Nguyen's account for Taylor's online information gathering. Trust plays a dual role for Taylor. First, she trusts a search engine to order search results according to epistemic relevance.³³ Taylor's disposition to immediately accept the search

²⁹ Nguyen, "Trust," 28, 44-45.

³⁰ Nguyen, "Trust," 44-45.

³¹ Nguyen, "Trust," 5.

³² A skeptic might protest that John would feel betrayal not towards the pacemaker, but towards the designers or doctors who implanted it. A similar objection could be levied against trust in search engines. If Taylor trusts the search engine, she will not feel betrayed by the search engine, but the search engine's designers. If the feeling of betrayal is directed towards the designers or implementers when objects fail, the skeptic argues, it is not obvious objects are proper targets of trust, after all.

Addressing this challenge—using Google Search's algorithm as example—Nguyen insists trust in designers and in the objects they create are distinct. Nguyen, "Trust," 8-9. To my understanding, this is related to the Nguyen's distinction between the participant stance and the integrative stance (see footnote 28 above). There is a distinction between the betrayal I would feel towards the designers *qua* moral participants and the object *qua* part of my agency. This is why I believe Nguyen maintains trust in an object can often be different from trust in the object's designer. Nguyen, "Trust," 8-9. In other words, John and Taylor can feel betrayal towards their pacemaker and search engine, respectively, insofar as the objects fail to integrate properly into their agency.

³³ Nguyen writes that we trust Google Search order search results according to relevance, and to guide our attention. Nguyen, "Trust," 20n4, 42. Note that I'm claiming Taylor trusts the search engine to order for *epistemic* relevance, specifically.

A skeptic might protest at the claim that Taylor trusts search engines in the latter sense. They might say: "Surely, Taylor does not trust the search engine. We seldomly feel betrayed (if at all) by search engines when they provide us with epistemically irrelevant information. We might feel annoyed—even very annoyed—but we do not feel betrayed."

There might be a plausible explanation for why we often do not feel betrayed. We often do not feel betrayed by search engines simply because we are not aware they provide us epistemically relevant results. (Thanks to Sebastian Watzl for this point.) As novices, we often turn to search engines because we do not know what is most epistemically relevant. Without the expertise to recognize epistemic irrelevant information, we do not recognize the search engine's failure. Hence, we would not feel betrayed. But would an expert feel betrayed

results as ordered according to epistemic relevance liberates time and mental energy. Instead of visiting webpages and finding a good expert haphazardly, she can use the search engine to direct her search more prudently. She can focus on assessing the experts the search engine suggests are most epistemically relevant for her, thereby increasing the likelihood of achieving her goal of finding a good expert about the COVID-19 vaccine.

Second, Taylor trusts experts to provide her with epistemically relevant information about the COVID-19 vaccine. Once an expert has been assessed as trustworthy, a robust disposition to immediately accept the information contained in an expert testimony as epistemically relevant is a prudent way to manage limited resources. She can trust the expert to have looked at the information available on the topic and assessed it for epistemic relevance. Therefore—ideally, at least—Taylor should be able to take the expert’s testimony as containing the most epistemically relevant information on the topic of the COVID-19 vaccine. Outsourcing her epistemic agency in this way frees Taylor to perform other tasks, such as finding the nearest vaccine center.

In sum, trusting search engines and experts helps Taylor find epistemically relevant information online in a prudent way—compensating for her lacking knowledge and her cognitive and practical limitations. Next, we will consider the capacity to direct trust and what an epistemic individualist conception would look like. Before doing so, though, I want to briefly address a worry.³⁴

One might worry that proceeding on an account of trust as outsourcing agency to the environment, I am stacking the cards in favor of the extended conception of the virtue of well-directed trust. By understanding trust as extending agency to the environment, the worry goes, I am smuggling in a metaphysical assumption of trust effectively guaranteeing the extended conception will be the most plausible one.

There are two reasons why the unquestioning attitude account ought not worry us in this respect. First, Nguyen stresses that his account is not wedded to any strong metaphysical

by their search engine if they noticed it provided epistemically irrelevant results about their field of expertise? The skeptic can still insist they would only feel very annoyed, but not betrayed. Interestingly, Nguyen opens up for other strong negative reactions counting as evidence of a trust relationship. Nguyen writes that “[...] failures [of agential integration] can ground sharply negative attitudes of betrayal, or *something very close to it* [my italics].” Nguyen, “Trust,” 38. So, I gather, provided the expert’s annoyance is sufficient strong enough to approximate a feeling of betrayal, Nguyen would still count it as evidence of a trust relationship with Google Search.

Whether Nguyen is correct that the relationship still counts as trust would take us too far afield. For the sake of discussion, I will assume both Taylor and the expert in this footnote count as trusting the search engine.

³⁴ Thanks to Sebastian Watzl for presenting me with this worry.

commitment about extension—such as the claim that trusting something makes it a part of the truster’s mind.³⁵ Second, even if trust implied strong metaphysical commitments, the *virtue* of well-directed trust can still be an individualist virtue. As I will argue below, an epistemic individualist can propose a conception of the virtue entirely in terms of individual factors—more specifically as a disposition of an agent’s capacity to direct trust. So, though I will ultimately argue the virtue of well-directed trust is best conceptualized as an extended virtue, I do not proceed on account of trust precluding plausible individualist conceptions.

1.2: The capacity to direct trust

We have determined trust as an unquestioning attitude whose role is to help an agent manage their limited cognitive and practical resources. This section will specify *how* an agent come to trust in the first place. Once we know how agents direct their trust, we can better determine how it could be done well. I assume agents direct trust using a capacity— ‘the capacity to direct trust’—which this section provides a model of.³⁶ Using this model, I will argue we can explain why Taylor directs her trust towards the WHO instead of the natural medicine expert.

My model of the capacity to direct trust has three components: inputs, a trustworthiness assessment process, and an output (an unquestioning attitude). My presentation will focus on the trustworthiness assessment process. This will be most pertinent component for determining the individualist conception of the virtue of well-directed trust.

Let us consider each component in turn, starting with the input. Once an agent or an object is considered a possible candidate for trust with respect to a given function, the truster takes in various pieces of information as input to the trustworthiness assessment process.³⁷ For example, when Taylor encounters the WHO in her search results, she starts by gathering and processing information about them. The search result headline, the search result snippet, the written testimony on the web page—all of this information (and probably more) is taken as input to the trustworthiness assessment process.

The trustworthiness assessment process is the second component of the model. The process aims to assess whether the agent or object is trustworthy with respect to a particular function.

³⁵ Nguyen, “Trust,” 27n7.

³⁶ The capacity to direct trust is similar to Anderson’s idea of a layperson’s “[...] second-order capacity to judge trustworthiness and consensus [...]” Anderson, “Democracy,” 145. Although trustworthiness judgements might be an intermediary step between trustworthiness assessment and the adoption of an unquestioning attitude, for simplicity, I have chosen to abstract away this intermediary step. My model is also inspired by Onora O’Neill emphasis on trustworthiness, as well as her notion “placing” and directing trust.” O’Neill, “Trust,” 5.

³⁷ The existence and specifics of the trustworthiness assessment process is an empirical question. Inputs suggest perception plays a role in the process. A treatment of these aspects is beyond the scope of my thesis.

Put differently, based on the inputs, the process estimates the reliability with which the agent or object can perform a particular function. For example, in the case of the WHO, the process uses, among other things, the information in the search result snippet to estimate how reliably the WHO will provide epistemically relevant information about the COVID-19 vaccine. Similarly, the trustworthiness assessment process can assess the trustworthiness of objects like search engines—at least in principle. By considering information such as the type of search results it produces, for instance, it could estimate the search engine’s reliability to provide epistemically relevant search results about the COVID-19 vaccine. However, as I will argue in a later section, assessing the trustworthiness of search engines will turn out to be challenging in practice. For now, though, the upshot is the trustworthiness assessment process is where the bulk of the work in the capacity to direct trust happens.

Further, I model the trustworthiness assessment process as working in two modes: automatic and reflective.³⁸ Before proceeding to describing the modes, let me briefly motivate why my model includes them in the first place. First off, plausibly, given her limited cognitive and practical resources, Taylor must be prudent with how she performs her trustworthiness assessments. According to Daniel Kahneman, dividing thinking between automatic and deliberative modes “[...] minimizes effort and optimizes performance.”³⁹ On this view, deliberative thinking often require effort.⁴⁰ Automatic thinking, by contrast, is quick and often effortless.⁴¹ While automatic thinking is more effective, deliberative thinking is activated when tasks are difficult, or something unexpected happens.⁴² Similarly, I suggest, the capacity to direct trust switches between automatic and reflective trustworthiness assessments. This is meant to ensure Taylor achieves her epistemic goals with limited cognitive and practical resources. Second, the distinction us explain phenomena related to epistemic character. For example, in Chapter 1’s sketch of WTD, I suggested an agent can get an impression or “gut feeling” about someone’s trustworthiness. The notion of automatic trustworthiness assessments lets us cash this out in more specific psychological terms. Finally, the notion of

³⁸ This is inspired by Gloria Origgi’s distinction between “default” and “vigilant” trust. Origgi, “Epistemic Injustice and Epistemic Trust,” 224. It is also inspired by Daniel Kahneman’s presentation of System 1 (automatic) and System 2 (effortful, conscious) thinking. Kahneman, *Thinking, Fast and Slow*, 20-25.

³⁹ Kahneman, *Thinking, Fast and Slow*, 25. Similarly, Relevance Theory claims human cognition tends to process the most relevant input (i.e., the input with greatest “positive cognitive effect” and least effortful). Wilson and Sperber, “Relevance Theory,” 608-10.

⁴⁰ Kahneman describes System 2 as effortful and deliberative. Kahneman, *Thinking, Fast and Slow*, 20-21.

⁴¹ Kahneman describes System 1 as operating automatically, with little effort. Kahneman, *Thinking, Fast and Slow*, 20.

⁴² Kahneman, *Thinking, Fast and Slow*, 24-25.

having automatic trustworthiness assessments also coheres well with Nguyen's claim that trust often happens in the "background": we do not notice it until it is endangered.⁴³

Let us start with the automatic mode of trustworthiness assessments. In the automatic mode, the assessment process uses various factors to assess a target as trustworthy without reflective deliberation. I suggest the capacity has a set of reliability criteria.⁴⁴ The precise criteria used by the capacity is an empirical question. However, for the sake of discussion, I will assume the criteria are stored as trust-related heuristics, biases, and background beliefs (e.g., the credentialing beliefs we saw in Chapter 1).⁴⁵ For example, there might be a trust-related heuristic that if an agent or object is trustworthy with respect to one function, chances are they are trustworthy on closely related functions. To illustrate, suppose Taylor has reflectively assessed WHO as a reliable source of epistemically relevant information about other health-related topics. She might then operate on a heuristic that the WHO will also be a good source of source on the topic of the COVID-19 vaccine. Of course, this is not a valid inference. It does not *necessarily* follow that the WHO will be trustworthy. Yet, in many situations, the heuristic will be helpful. Using heuristics, biases and background beliefs like these, I imagine the capacity to direct trust makes automatic trustworthiness assessments.

Having seen how it works in its automatic mode, let us see how the assessment works in the reflective mode. In this mode, considerations about the agent or object's trustworthiness will enter into the agent's conscious awareness. To illustrate, suppose Taylor was not familiar with the WHO. Before she decides to adopt an unquestioning attitude, she deliberates about whether the WHO can perform the function of being a reliable source about the COVID-19 vaccine. She will probably reflect upon the language used, whether there are any scholarly references, as well as looking up the WHO's credentials on Wikipedia.⁴⁶ In this way, the trustworthiness assessment process can also happen in a reflective mode.

⁴³ Nguyen, "Trust," 14-15.

⁴⁴ For an overview and discussion of criteria for trustworthiness assessment of scientific expertise by lay persons with access to the Web, see Anderson, "Democracy," 146-49. See also Origgi's related discussion about various mechanisms that make us trust. Origgi, "Epistemic Injustice and Epistemic Trust," 227-33.

⁴⁵ This point is inspired by several sources. It is inspired by the role of "background beliefs" in Sperber et al.'s account of epistemic vigilance. See especially Sperber et al., "Epistemic Vigilance," 374-76. It is also inspired by the role of trust heuristics in Gloria Origgi's presentation of mechanisms making us trust. Origgi, "Epistemic Injustice and Epistemic Trust," 227-33. Lastly, it is also inspired by Gabrielle M. Johnson's account of bias as a "necessary response" to the challenge that we always access the world via limited empirical evidence, supporting several theories and conclusions. Johnson, "Structure of Bias," 1195-96.

⁴⁶ I am drawing upon Elizabeth Anderson's discussion of how laypersons can use Wikipedia to make trustworthiness judgements about the credentials of global warming experts. Anderson, "Democracy," 149-53.

How does Taylor navigate between the two modes to perform well without spending too much effort? As this is an empirical question, I will content myself with providing a few tentative suggestions. One suggestion is that she will tend to engage in reflective trustworthiness assessments in unfamiliar or surprising situations.⁴⁷ Another suggestion is that Taylor navigates the two modes depending on the stakes involved in adopting an unquestioning attitude.⁴⁸ The agent will—or at least *should*—engage in more deliberative thinking about trustworthiness if trusting something has high stakes.⁴⁹ Finding out whom to trust about the COVID-19 vaccine arguably has high stakes. So, although Taylor has a heuristic to automatically trust the WHO, the stakes will make her engage in a reflective trustworthiness assessment.

Let us turn to the final component of the model, the output, and explain Taylor's trust pattern. The capacity outputs an unquestioning attitude, provided the source of the testimony is considered as meeting a certain level of trustworthiness with respect to the assessed function of the target. Taylor assesses—automatically or reflectively—the WHO to be a reliable source of information about the COVID-19 vaccine. Therefore, she adopts an unquestioning attitude towards the WHO as a source of information on that topic. For opposite reasons, Taylor does not adopt an unquestioning attitude towards the natural medicine expert. Analogously, if Taylor makes an automatic or reflective assessment of, say, Google Search as trustworthy as ordering epistemically relevant search results about the COVID-19 vaccine, the capacity will produce an unquestioning attitude.

1.3: What an epistemic individualist conception of the virtue of well-directed trust should look like

With an understanding of trust as unquestioning attitude and a model of the capacity to direct trust, we will now turn to what the individualist conception of the virtue of well-directed should look like to withstand the structural critique in Chapter 1. I will argue the individualist conception should be a disposition of the capacity to direct trust, more specifically a disposition to adopt unquestioning attitudes in a truth-conducive and error-minimizing way.⁵⁰

⁴⁷ Compare to Kahneman's comment that System 2 (effortful thinking) is activated when something unexpected happens. Cf. Kahneman, *Thinking, Fast and Slow*, 24.

⁴⁸ This point is inspired by Gloria Origgi's comments on the responsibility of to be vigilant about our trust in high-stakes situations. Origgi, "Epistemic Injustice and Epistemic Trust," 224, 227.

⁴⁹ Origgi, "Epistemic Injustice and Epistemic Trust," 224, 227. See previous footnote.

⁵⁰ I follow Heersmink definition of virtues as truth-conducive and error-minimizing. Heersmink, "Virtue Epistemology," 3. For other trust-related epistemic virtues proposed in the philosophical literature, see Chapter 1, footnote 100. In contrast to these, the individualist conception of the virtue of well-directed trust on offer here is couched terms of Nguyen's unquestioning attitude account of trust (see section 1.1).

On this view, the virtue of well-directed trust is a mean-virtue between, on the one hand, the vice of epistemic gullibility and, on the other hand, the vice of epistemic hypervigilance.

To motivate what it means to trust well, I want to distinguish between two types of justified trust: rational and well-directed trust.⁵¹ Carolyn McLeod writes that trust is warranted when it is rational—such as when it is supported by good evidence.⁵² Suppose Michael has good evidence for trusting the door opener at the tram. He has never experienced a malfunctioning tram door opener in the city. Moreover, he knows the city prioritizes public transportation highly. So, his trust in the tram door opener is rational because he has good evidence and reasons to trust it will open.

Trust can also be justified when it is well-directed. I will understand trust as being “well-directed” when it is directed towards a trustworthy agent or object.⁵³ Michael’s trust in the tram door opener is well-directed, if the tram door opener is in fact trustworthy with respect to opening the door. In other words, it *reliably* opens the door when someone pushes the opener. However, trust can be rational without being well-directed.⁵⁴ Suppose Michael, upon trying to exit the tram, discovers the door opener is malfunctioning. Insofar as he had good reasons and evidence, his trust was rational. Yet, because the door opener turned out not to be trustworthy, his trust was *not* well-directed. In this way, trust can be rational without being well-directed. Throughout this chapter, I will mostly focus on well-directed trust. However, when I discuss trusting potentially biased search engines, I will briefly return to the topic of rational trust.

Understanding what it means for trust to be well-directed, let us turn to formulating the individualist conception of the virtue of well-directed trust. As mentioned, I will argue the conception ought to be formulated disposition to adopt an unquestioning attitude in a truth-conducive and error-minimizing way. I also mentioned that I will argue for the individualist conceptualization being a mean-virtue. That is to say, the virtuous disposition is a virtuous middle between two distinct vices, which I call “epistemic gullibility” and “epistemic

⁵¹ I am inspired by Carolyn McLeod’s distinction between two forms of “warranted trust”: “justified” (trusting rationally, such as “on good evidence”) and “well-grounded” (trusting trustworthy persons). McLeod, “Trust.”

⁵² McLeod, “Trust.”. See also previous footnote.

⁵³ Onora O’Neill also distinguishes between “well-directed” and “misdirected” trust, and I follow her in stressing the importance of trust being directed towards trustworthy things. O’Neill, “Trust,” 5. Similarly, the notion of directing trust well or badly takes inspiration from Nguyen’s suggestive term: “badly aimed trust”. Nguyen, “Cognitive islands,” 2817. My definition is also related, but importantly different from what Carolyn MacLeod calls “well-grounded trust” (see footnote 51 above). McLeod’s term only concerns trust in agents. McLeod, “Trust.” But following Nguyen’s analysis of trust as unquestioning attitude, presented in section 1.1, my notion of well-directedness also includes trusting trustworthy objects.

⁵⁴ McLeod makes a similar point when she writes that trust can from time to time be justified, without the trusted person being trustworthy. McLeod, “Trust.”

hypervigilance.” To get a better sense of what the virtue of well-directed trust is, let us start by describing the vicious dispositions first.

Epistemic gullibility is a disposition of the capacity to direct trust to adopt an unquestioning attitude without sufficient trustworthiness assessment. In other words, the epistemically gullible agent tends to adopt an unquestioning attitude when they ought not.⁵⁵ Epistemic gullibility manifests in at least two ways. First, epistemic gullibility could make an agent fail to engage in reflective trustworthiness assessment when they ought to. An epistemically gullible agent might fail to appreciate the stakes involved in taking the COVID-19 vaccine, leading them to fail to reflectively assess the credentials of purported experts about the vaccine. While this could lead to a positive outcome if the expert happens to be trustworthy, it could also make the agent liable to trust bad experts. Consequently, the epistemically gullible agent is vulnerable to immediately accept false beliefs about the COVID-19 vaccine.

A second manifestation of epistemic gullibility is to be badly disposed with respect to reflective trustworthiness assessments. When the epistemically gullible agent engages in deliberations about trustworthiness, their reasoning might fail to focus on what is most relevant with respect to trustworthiness. Instead of focusing their reasoning on an expert’s credentials, say, they might focus on how well-articulate the expert is. This also makes them disposed to adopt an unquestioning attitude when they ought not to. Clearly, directing trust in these epistemically gullible ways will not be truth-conducive and error-minimizing.

Epistemic hypervigilance is the opposite extreme of epistemic gullibility. Epistemic hypervigilance is a disposition to excessively assess an agent or object’s trustworthiness. In other words, the epistemically hypervigilant person fails to adopt an unquestioning attitude towards an agent or object when they ought to.⁵⁶ A manifestation of this disposition could be an agent getting stuck in the deliberation phase due to an excessively high threshold for trustworthiness. Steve, a novice about vaccines, is worried the recommendations the WHO

⁵⁵ This disposition echoes Nguyen’s understanding of gullibility as adopting an unquestioning attitude too readily. Nguyen, “Trust,” 43. In this paper, Nguyen does not formulate gullibility as an epistemic vice. Upon finalizing the thesis, though, I learned that, in a different paper, he lists gullibility as an example of an epistemic vice. Nguyen, “Seductions of Clarity,” 230.

⁵⁶ This notion is inspired by work on (epistemic) vigilance by Gloria Origgi and Sperber et al. Origgi, “Epistemic Injustice and Epistemic Trust.”; Sperber et al., “Epistemic Vigilance.” Moreover, in situations perceived as fearful and uncertain, Dyrendal and Emberland write, individuals tend to seek information to make sense of their situation. Dyrendal and Emberland, *Hva er konspirasjonsteorier*, 46. Fear and uncertainty make individuals exhibit vigilance (Nor. “vaktksomhet”), and they can become excessively concerned with signs of danger. Dyrendal and Emberland, *Hva er konspirasjonsteorier*, 46. This can lead them to engage in conspiratorial thinking. Dyrendal and Emberland, *Hva er konspirasjonsteorier*, 46-47.

gives today might turn out to have been wrong six months later. Instead of immediately accepting what they say, he scrutinizes their every recommendation for trustworthiness. The problem in Steve's case is the WHO will never live up to his trustworthiness threshold. More likely, all things considered, he would probably be epistemically better off by simply trusting the WHO. Similarly, as I will discuss later, an agent might be epistemically hypervigilant if they do not trust a search engine, despite the possibility of the search engine being biased.

Now that we understand the vices of epistemic gullibility and epistemic hypervigilance, we can now consider the virtue of well-directed trust as a mean virtue between the vices.⁵⁷ Assuming the virtue of well-directed trust is best conceptualized in this manner, how would Taylor differ from the vicious agents? Unlike the epistemically gullible person, Taylor will recognize the stakes associated with the COVID-19 vaccine and engage in appropriate degree of reflective trustworthiness assessment. And unlike Steve, Taylor will not engage in excessive trustworthiness assessments. She will have a realistic threshold for trustworthiness, meaning she is willing to accept some risk the trustee might not be perfectly reliable. Once the candidate has been assessed as sufficiently trustworthy, Taylor does not hesitate about adopting an unquestioning attitude. For example, she will assess the WHO as sufficiently trustworthy and thereby trust them. This disposition is likely to be truth-conducive and error-minimizing: it reduces the risk of forming false beliefs by trusting too easily, while it increases the likelihood of gaining true beliefs by trusting sources deemed sufficiently trustworthy.

In this way, the epistemic individualist can argue the virtue of well-directed trust is plausibly conceptualized as an individualist virtue. A virtuous disposition of an agent's capacity to direct trust accounts for the virtue of well-directed trust purely in terms of the individual. Yet, in the next section, I will argue the virtue of well-directed trust is best conceptualized as an extended virtue.

2: The Virtue of Well-Directed Trust Is Most Plausibly an Extended Virtue

In the previous section, I argued how I think an individualist conception of the virtue of well-directed trust would have to look like to meet the structural critique from Chapter 1. In this chapter, though, I will argue that the virtue is better conceptualized as an *extended* virtue. In brief, I argue Taylor can possess the disposition identified with the individualist conception

⁵⁷ This definition therefore echoes Heersmink's definition of an epistemic virtue as a mean between two vices. Heersmink, "Virtue Epistemology," 5.

without directing her trust well, due to the search engine’s bias towards epistemic irrelevance. Because the search engine’s bias exerts considerable influence upon Taylor’s trust direction, I will maintain the virtue of well-directed trust is best conceptualized as an extended virtue.

To manage her cognitive and practical limitations, Taylor submits queries to search engines like “Should I take the COVID-19 vaccine?”. The search engine is meant to do this by providing Taylor with an ordering of the most epistemically relevant websites about the COVID-19 vaccine. Recall Taylor has two tasks to achieve her goal of learning about the COVID-19 vaccine using search engines. Her first task is to find expert testimony and her second is to assess the trustworthiness of the purported expert. If she considers the expert trustworthy, she will trust—i.e., adopt an unquestioning attitude towards—the expert as a source epistemically relevant information about the COVID-19 vaccine.

Despite offloading cognitive and practical load to search engines, Taylor still faces cognitive and practical limitations. There will likely be thousands, if not millions, of potentially relevant search results from purported experts on the COVID-19 vaccine. Clearly, Taylor do not have the time or mental energy to assess every search result, be it automatically or reflectively.⁵⁸ So, it is only epistemically practicable for Taylor to assess a subset of purported experts in her search results. Call this the “practicable assessment subset”. If Taylor only has the resources to assess the trustworthiness of the expert in the search results from the few search result pages, these search results will be the members of her practicable assessment subset. Thus, Taylor’s cognitive and practical limitations restricts her trustworthiness assessments—and thereby trust—to the experts in her practicable assessment subset.

Crucially, whether she finds a good expert to trust using the search engine depends on whether a good expert is a member of the practicable assessment subset. Suppose the WHO’s website is not among Taylor’s practicable assessment subset. As a result, she will not assess the WHO’s trustworthiness *qua* source of epistemically relevant information about the COVID-19 vaccine. Recall that a trustworthiness assessment—automatic or reflective—is necessary condition for trust. So, because she cannot assess the WHO is, she will not adopt an unquestioning attitude towards the WHO. How well Taylor directs her trust, then, depends on the extent to which the practicable assessment subset contains good experts.

⁵⁸ The Google search query “should I take the covid vaccine?”, submitted by me in Oslo on 15.02.2022, resulted in 587,000,000 results.

In turn, the number of good experts in the practicable assessment subset depends on the search engine's bias towards epistemic irrelevance. Suppose the search engine Taylor uses employs a ranking algorithm biased towards epistemic irrelevance. The ranking algorithm might be biased because epistemically irrelevant search results are more profitable for the search engine owners. For instance, presenting epistemically irrelevant search results might generate more data the owners can use to make their targeted advertisement more accurate.⁵⁹ As a result, Taylor might find the natural medicine expert from chapter 1 in her search results. The upshot is: the *more* biased the search engine is towards epistemic irrelevance, the *greater* the probability the practicable assessment subset contains a *bad* expert like the natural medicine expert, and the *lower* the probability it will contain a *good* expert like the WHO.

Note that how well Taylor directs her trust is not merely a matter of her individual disposition—it is also a matter of how biased the search engine is. My argument suggests two causal contributions to the wellness of Taylor's trust: Taylor's individual disposition to direct trust well and the online epistemic tool's bias towards epistemic irrelevance. As my line of argumentation suggests, the presence of good experts in the practicable assessment subset is crucial for Taylor directing her trust well. The search engine's bias towards epistemic irrelevance is therefore an important causal contribution in how well Taylor directs her trust. Thus, if the search engine is greatly biased towards epistemic irrelevance, Taylor can have the disposition of the individualist conception of well-directed trust *without* directing her trust well.

If so, this suggests having the individualist virtue of well-directed trust is not sufficient for directing trust well. Suppose Taylor adopted a strategy of trusting the most trustworthy expert among the practicable assessment subset. This seems like a reasonable—or at least a justifiable—strategy. After all, she must make a decision at some point, with the information she has, about whether to take the COVID-19 vaccine.

To see why search engine bias is a problem for the individualist conception of the virtue of well-directed trust, consider two scenarios. First, consider a scenario where the search engine is moderately biased towards epistemic irrelevance. The likelihood is moderately high the

⁵⁹ This point is inspired by Shoshana Zuboff's summary of the "logic and operations of surveillance capitalism," where Google Search's targeted advertisement features as an example. Zuboff, *Age of Surveillance Capitalism*, 93-97. Zuboff does not frame her analysis in terms of epistemic relevance. Yet, I believe her economic analysis is compatible with it. Although I assume the ranking algorithm of Taylor's search engine is affected by economic interests, note that Google is adamant that they do not let payments affect their ranking. Google, "Our Approach – How Google Search Works."

practicable assessment subset contains mediocre or bad experts. Seeing as Taylor is a novice about the COVID-19 vaccine, this raises the probability that Taylor will not come to trust a good expert. But it is still not impossible a good expert is among her subset. Next, consider the worst-case scenario: the search engine is severely biased towards epistemic irrelevance. The risk of trusting a bad or mediocre expert is even more pronounced. The probability the practicable assessment subset contains a good expert is quite low. These two scenarios suggest Taylor's disposition *alone* does not make her direct trust in a truth-conducive, error-minimizing manner. This is a puzzling result for the individualist conception, insofar as the disposition was meant to ensure the agent directs their trust well.

The influence of the search engine's bias suggests it is more appropriate to conceive of well-directed trust as an extended virtue. By 'extended virtue', I mean a virtue where at least one considerable cause of the virtuous activity's result is outside the individual.⁶⁰ For example, the virtue of well-directed trust would be extended if the search engine's bias is a considerable cause of whether Taylor winds up trusting a good expert. By 'considerable cause,' I appeal to an intuitive distinction between mere background conditions and more "substantial" causes. As I interpret him, James Woodward suggests one can defend a metaphysical distinction between background conditions and causes.⁶¹ So, I will assume there is a metaphysical distinction between background conditions and what I call "considerable causes."

One might object that a reliable internet connection, in a sense, also is an external cause of the wellness of Taylor's trust direction.⁶² True, but I find it more plausible to think of an internet connection as a *background condition* for the result of her virtuous activity. Seeing as the exercise of most virtues requires meeting *some* background conditions, we risk making the set

⁶⁰ Although I do not make any metaphysical commitment to a version of the extended mind thesis, my extended virtue definition draws inspiration from this literature—especially its emphasis on tools. Clark and Chalmers, for instance, view language as a tool for extended cognition. Clark and Chalmers, "Extended Mind," 18. For a discussion combining versions of the extended mind thesis, trust, tools and Heersmink's virtue responsibilism for the Internet (without adding a virtue of well-directed trust), see Schwengerer, "Online Intellectual Virtues."

⁶¹ This interpretation is based on Woodward's suggestion that properties of causes can capture the distinction between causes and background conditions. Woodward, "Causation in biology," 317, 317n24. However, I do not intend to commit myself to Woodward's interventionist framework being the best framework for causality. For a brief introduction to the framework, see Woodward, "Causation in biology," 289-91. The comprehensiveness of bringing theories of causality into the discussion is beyond the scope of this thesis.

⁶² Heersmink seems to be aware of a related discussion about distribution of epistemic responsibility not only to Internet users, but also to governments, designers of search engines, and more. Heersmink, "Virtue Epistemology," 8. In this discussion, Heersmink seems to think the primary epistemic responsibility nevertheless lies on Internet users to acquire epistemic virtues. I disagree with this position, because I think search engines exert a considerable influence upon Internet users' virtuous behavior.

of extended virtues bloated. Thus, I want to reserve ‘extended virtue’ only for virtues where the external cause exerts a considerable influence on the virtuous activity’s result.

A skeptic might not be convinced that I have made the considerable influence of the search engine’s bias sufficiently plausible. They might object that I have only demonstrated the search engine’s bias to be nothing more than a background condition—just like access to the Internet—for the wellness of Taylor’s trust direction. A satisfactory demonstration of the search engine’s bias being a considerable cause would require comprehensive engagement with theories of causality. This is beyond the scope of this chapter and thesis. However, I will briefly explain why I think the influence counts as a considerable cause.

Woodward suggests one could use causal properties—e.g., specificity—to differentiate between mere background conditions and causes.⁶³ One way to understand “specificity” is the property of a cause to “[...] modulate the state of [the effect] in a fine-grained way.”⁶⁴ Based on this observation, I suspect one can argue that the search engine’s bias modulates the wellness of Taylor’s trust in a fine-grained manner by including more or less epistemically relevant experts in her practicable assessment subset. If my suspicion is borne out by a more comprehensive demonstration, I believe it is warranted to consider the causal influence of the search engine’s bias as more than a mere background condition.

Summing up, I claim my line of argumentation has rendered plausible the claim that the search engine’s bias, via the practicable assessment subset, exerts a considerable influence upon how well Taylor’s directs her trust. Therefore, I conclude the search engine’s bias towards epistemic irrelevance qualifies as a considerable cause of the wellness of Taylor’s trust direction. Since the search engine’s bias is an external cause, the virtue of well-directed trust satisfies the definition of an extended virtue. Therefore, I maintain my argument has shown the virtue of well-directed trust is best conceptualized as an extended, not an individualist, virtue.

If my argument holds—which I believe it does—the consequences would be dire for the epistemic individualist from Chapter 1. Insofar as the virtue of well-directed trust is best

⁶³ Woodward, "Causation in biology," 317, 17n24. According to Woodward suggestive comments, other causal properties like stability could also differentiate a cause from a background condition. As I read him, though, specificity alone could make a factor a cause. He hypothesizes that a causal factor *X* is considered a background condition of the effect *Y* when “[...] *X* is a non-specific cause of *Y* and/or [my italics] the *X*–*Y* relationship is unstable [...]” Woodward, "Causation in biology," 317. Conversely, I take it, *X* would be considered a cause of *Y*, if it was specific and/or stable. In other words, specificity alone is sufficient for a causal factor to count as a *bona fide* cause.

⁶⁴ Woodward, "Causation in biology," 305.

conceptualized as an extended virtue, the epistemic individualist cannot appeal to this virtue in defense of the structural critique raised in Chapter 1.

Before I conclude this chapter, though, I want to address an objection against my argument.

3: The Virtue of Well-Directed Trust Not Extended, After All?

3.1: Objection: Taylor was epistemically gullible for trusting the search engine

An epistemic individualist might resist my conclusion about the virtue of well-directed trust being best conceptualized as an extended virtue. The epistemic individualist could resist my argument by arguing Taylor was epistemically gullible for trusting a potentially biased search engine in the first place. Here is how they make their objection.

The epistemic individualist objects that my argument proceeds on a problematic assumption: Taylor's trust in the search engine is compatible with the disposition of well-directed trust. This is a problematic assumption, according to the epistemic individualist. My argument already leverages the fact that search engines are potentially biased towards epistemic irrelevance. Facing the possibility of a potentially biased search engine, the epistemic individualist argues, it would be odd for Taylor to adopt an *unquestioning* attitude. It would be odd, that is, to trust the search engine to provide epistemically relevant search results about the COVID-19 vaccine. Taylor's trust in the potentially biased search engine is evidence of an epistemically gullible disposition, the epistemic individualist contends.

To see why, the epistemic individualist argues Taylor assessing the search engine as trustworthy—be it automatically or reflectively—is evidence of epistemic gullibility. To start off, the objector asks us to suppose Taylor's trust in the search engine was the result of an automatic assessment. Information about the possibility of search engine bias towards epistemic irrelevance has become widely known under names such as “filter bubbles” and “personalization.”⁶⁵ Thus, Taylor is surely aware of the possibility that the search engines she uses are potentially biased towards epistemic irrelevance. Otherwise, we would suspect Taylor did not possess the virtue of intellectual thoroughness, which we assumed she possesses.

Moreover, insofar as search engines play a crucial role in information gathering online, the

⁶⁵ I cite the following as evidence for search engine bias towards having become a commonplace. First, large media outlets have written about it. For example, media outlets have written about search engine bias, filter bubbles, and misinformation. See for example BBC, "How algorithms and filter bubbles decide what we see on social media - BBC Bitesize."; Berlatsky, "Google search algorithms are not impartial." Second, Bergstrom and West's non-technical book warns about the consequences of algorithmic personalization on social media. See especially Bergstrom and West, *Calling Bullshit*, 28.

stakes are high for trusting a potentially biased search engine.⁶⁶ So, if Taylor were disposed to trust well, the epistemic individualist insists, the information about potential search engine bias should have led her to recognize the stakes. Therefore, the epistemic individualist concludes, trusting based on an automatic trustworthiness assessment evidence of Taylor's epistemic gullibility.

Likewise, the epistemic individualist argues, Taylor's coming to trust trusting the search engine as a result of a reflective trustworthiness assessment is also evidence of an epistemically gullible disposition. Suppose Taylor *did* recognize the stakes, performed a reflective trustworthiness assessment, and trusted the search engine. The epistemic individualist maintains this would be evidence of Taylor's epistemic gullibility. After all, she has information questioning the search engine's ability provide epistemically relevant search results. If she trusts the search engine despite this information, the epistemic individualist believes this is evidence Taylor is not disposed to reflectively assess trustworthiness.

However, there are two challenges for the epistemic individualist's claim that Taylor can perform a reflective trustworthiness assessment. First, there is the challenge of untransparent search engines. It is not obvious how she can assess trustworthiness when information about the internal operations of search engines is not available to the public.⁶⁷ Second, there is the challenge of how Taylor, a novice about computer science, can make an adequate reflective trustworthiness assessment of search engines in the first place. Yet, I want to show these challenges are not necessarily decisive blows against the epistemic individualist objection. The epistemic individualist can argue neither transparent search engines nor expertise are necessary for Taylor to make a reflective a trustworthiness assessment of a search engines.⁶⁸

Recall our assumption that search engines rank search results with a ranking algorithm, more specifically a machine learning algorithm. In this case, we assume it is an algorithm "training" on data to improve its predictions of the search results a user will find epistemically relevant.

⁶⁶ See also Nguyen's discussion of epistemic gullibility, trust in Google Search, and the trusting agent's vulnerability to changes happening behind the scenes. Nguyen, "Trust," 42-43.

⁶⁷ Frank Pasquale notes lacking transparency of search engines like Google Search as a problem; though, he advocates for intelligibility, not mere transparency, as the solution. Pasquale, *The Black Box Society*, 7-8.

Also, after writing this chapter, the search engine Google Search has become more transparent. Google has released more information about how Google Search works. Google, "In-Depth Guide to How Google Search Works." Even though Taylor would have more information available in her assessment, the rest of this section will show the epistemic individualist objection does not require transparent search engines.

⁶⁸ The argument below is an adaptation of arguments made by Bergstrom and West. For an overview of the general argument, see Bergstrom and West, *Calling Bullshit*, 42-44. For a second overview of the argument applied to machine learning algorithms in particular, see Bergstrom and West, *Calling Bullshit*, 182-83, 192-93.

Carl T. Bergstrom and Jevin D. West argue one often does not need extended knowledge about computer science or statistics to assess a machine learning system.⁶⁹ Without inspecting the details of the algorithm, they argue one can evaluate it by focusing on its training data and its output.⁷⁰ If the machine learning algorithm is trained on biased data, the authors add, the technical specifics of *how* the prediction is made is often irrelevant.⁷¹ The reason is problematic training data will produce problematic predictions.⁷² Pertinent questions to ask about training data are, among others, how representative or relevant they are, their origins, and who made them.⁷³ Likewise, if the algorithm produces obviously implausible output, then the technicalities of how the output was produced often does not matter.⁷⁴ Let us illustrate the argument with an example.

Suppose Taylor is considering whether to trust her new digital thermostat to provide her with an accurate weather report for the Oslo region. The thermostat employs the newest machine learning technology to predict the weather. Taylor, we remember, has no computer science expertise. But she nonetheless assesses the thermostat's trustworthiness. First, she checks the training data. As the thermostat employs open-source code, Taylor can go to the manufacturer's website and find the data training data. She discovers that the machine learning algorithm is trained on weather data taken exclusively from California. As a result, she suspects the thermostat will not generalize well to colder areas like Oslo. Second, her suspicion is corroborated the thermostat's output. The thermostat shows 25 degrees Celsius, when there is snow outside. Clearly, the thermostat fails to perform its function. The upshot is: Taylor can reflectively assess the thermostat's trustworthiness *qua* reliable indicator of the outside temperature without being a computer science expert. And adopting an unquestioning attitude towards would be epistemically gullible.

The epistemic individualist could say Taylor could adopt a similar strategy for reflectively assessing a ranking algorithm's trustworthiness with respect to providing epistemically relevant search results. Suppose Taylor submits a search query about Latin, which she knows a lot about. First, just like Taylor did for the thermostat, Taylor could assess whether the

⁶⁹ Bergstrom and West, *Calling Bullshit*, 42-43, 183.

⁷⁰ Bergstrom and West, *Calling Bullshit*, 43.

⁷¹ Bergstrom and West, *Calling Bullshit*, 43, 183.

⁷² Bergstrom and West, *Calling Bullshit*, 183. As the authors note here, bad labelling of data also contributes to making predictions bad. To avoid needless complications, I only consider training data in my argument.

⁷³ Bergstrom and West, *Calling Bullshit*, 43, 192.

⁷⁴ Bergstrom and West, *Calling Bullshit*, 43. The authors make this point about black boxes in general. So, it should also apply to machine learning algorithm.

ranking algorithm's training data is skewed towards epistemic relevance with respect to Latin. Second, she could also check the output of the ranking algorithm. If Taylor finds mediocre or bad experts on Latin in the practicable subset, she could take this as evidence the search engine's ranking algorithm being biased towards epistemic irrelevance. Thus, by adopting this strategy, the epistemic individualist maintains it is epistemically practicable for Taylor to reflectively assess the search engine's trustworthiness—all without transparent search engines and without computer science expertise.

Summing up the objection, the epistemic individualist claims my argument does not follow because it proceeds on a problematic assumption: Taylor's trust in a (potentially) biased search engine is compatible with the individualist conception of well-directed trust. On the contrary, the objector argues, Taylor's trust is evidence of her epistemic gullibility. That search engines are biased is a plausible prospect. So, the objector argues, Taylor would be epistemically gullible to assess the search engine as trustworthy—be it automatically or reflectively. The objector concludes the extended virtue argument rests on a problematic assumption.

Next, I will argue that the objection fails because Taylor is not necessarily as epistemically gullible as the objection makes her out to be. Moreover, even if she were, it would not show the individualist conception of the virtue of well-directed trust as the most plausible.

3.2: Reply

3.2.1: Taylor is not necessarily as epistemically gullible as the objection claims

I will present two arguments why Taylor is not necessarily as epistemically gullible as the objector make her out to be.

To start off the first argument, recall the objector claimed Taylor's was epistemic gullibility because it was epistemically practicable to assess the search engine's trustworthiness with the proposed strategy. The strategy, we recall, consists of scrutinizing the search engine's training data and search results in a knowledge domain known to Taylor. This strategy will not work for Taylor. To start off, it is implausible Taylor can assess the search engine's training data for bias towards epistemic irrelevance. Finding training data for search engines can be hard to do.⁷⁵ To my knowledge, Google Search, for example, have not made their training data

⁷⁵ While writing this chapter, I conducted several search queries in an attempt to obtain training data for Google Search. My search query failed to find training data.

openly available.⁷⁶ Yet, even if Taylor had the training data, it is not obvious to me she could easily uncover bias towards epistemic irrelevance. A sophisticated online epistemic tool like a search engine likely requires a large amount of training data. After all, the number of topics the search engine is meant to provide relevant results about is very large.

To be sure, perhaps Taylor could employ software to aid her in assessing the training data.⁷⁷ By having statistical patterns described or visualized, one might suggest, Taylor might be able to uncover epistemic relevance. But this suggestion leads to a similar problem: To not be epistemically gullible, Taylor should assess the trustworthiness of software helping her assess the training data, as well. At this point, though, we are stretching what is reasonable to expect from a novice about computer science and statistics. It is more plausible to think Taylor could not assess a search engine's training data—even if she did have access to them.

Likewise, scrutinizing the search engine's output will not help Taylor make a good trustworthiness assessment. Just because the search engine provides epistemically relevant search results about Latin, it does not follow it will provide epistemically relevant search results about the COVID-19 vaccine.⁷⁸ Suppose the search engine happens to be trustworthy about Latin, but not about the COVID-19 vaccine. Taylor would be wrong to infer the search engine is trustworthy about the COVID-19 vaccine, even if it provides good experts about Latin in her practicable assessment subset. Even in the best-case scenario, where the search engine is trustworthy both about Latin and the COVID-19 vaccine, her inference would not be justified. Thus, it is not plausible Taylor can use the search engine's output about knowledge domains she is familiar with to assess its trustworthiness about knowledge domains she is not familiar with.

⁷⁶ There could be several reasons why training data is not openly available. The following reasons are speculations on my part. There might be technical difficulties in making all of these data available. There might also be economic reasons against transparency. For one, the company might risk losing their competitive advantage if someone uses the training data to create a better product. For another, as Odd Erik Gundersen noted to me in our conversation, transparency increases the risk of uncovering bias in their products could create a PR problem. Having training data open to the public—letting users like Taylor potentially discover bias towards epistemic irrelevance—might not be in the interest of Google. Thanks to Odd Erik Gundersen for a helpful discussion regarding the economic incentives involved in transparency.

⁷⁷ For example, pandas, a library for the programming language Python, lets one produce descriptive statistics for large data sets. NumFOCUS, "pandas.DataFrame.describe — pandas 1.5.3 documentation."

⁷⁸ Here I am adapting a structurally similar point made by Sebastian Watzl in our discussion of the epistemic individualist's strategy for trustworthiness assessment of search engines.

It is implausible the epistemic individualist's strategy will make Taylor capable of assessing the trustworthiness of the search engine. Thus, it is not as obvious Taylor's failure to assess the search engine's trustworthiness is the result of an epistemically gullible disposition.

I turn now to my second argument. It will argue that Taylor's trust in a *potentially* biased search engine is not evidence of an epistemically gullible disposition. While it is certainly desirable to avoid trusting experts and online epistemic tools one suspects are biased, there will very often exist uncertainty with respect to bias. There is always a possibility the expert or online epistemic tool is biased in one way or another. For a novice, it can be hard to assess for biases.

Still, a novice's trust can be justified insofar as it is rational. Recall from section 1.3 that trust can be justified by good reasons and evidence, despite not being well-directed. So, despite the possibility of bias, the novice's trust in the search engine might still be justified by good reasons and evidence. Suppose, for instance, Taylor reads that recent empirical literature suggests search engines are less biased than many have believed earlier.⁷⁹ In and of itself, this may not be sufficient for Taylor to assess her particular search engine as a trustworthy provider of epistemically relevant search results about the COVID-19 vaccine. Yet, these findings can be part of Taylor's *risk assessment* of trusting a potentially biased search engine—that is, whether the epistemic benefits of trusting the potentially biased search engine would outweigh the costs.

Whether it is rational to take the risk of trusting the potentially biased search engine will of course depend on the concrete epistemic benefits and costs in question. The point I want to make is Taylor's decision to take the risk is not necessarily a sign of epistemic gullibility. On the contrary, it can be a sign that she is disposed to reason well about trust under uncertainty.

Moreover, a reluctance to trust under certainty can be evidence of the vice of epistemic hypervigilance. If Taylor is reluctant to trust search engines due to potential bias, she should also consider the epistemic losses she incurs by not trusting. By not trusting, she might also forego the possibility of navigating the Internet more effectively. Therefore, one can make a case that Taylor would be epistemically hypervigilant not to do so. I am not saying the epistemic individualist objector claims an agent can accept *no* risk about search engine bias.

⁷⁹ For a brief discussion of the nascent empirical literature on the impact of ranking algorithms, where some doubt the extent to which personalized algorithms restrict access to diverse information, see Barberá, "Social Media," 41-44. Arguedas et al. gives an empirical overview, and argue search engines and social media is associated more varied news consumption. Arguedas et al., "Echo chambers," 17-20.

But this discussion does raise an interesting question about the degree of risk an agent well-disposed to reason about trustworthiness and should accept.

Summing up, I hope the second argument has showed Taylor need not be epistemically gullible for trusting a potentially biased search engine—contrary to what the epistemic individualist objection seemed to suggest.

3.2.2: Alternatives to trust do not suggest the individualist conception is most plausible

Despite my arguments to the contrary in the previous section, the epistemic individualist might not be convinced. They might still insist Taylor's trust in the search engine is evidence of epistemic gullibility. If this were true, the objector ought to propose an alternative for Taylor to find good experts about the COVID-19 vaccine online. I consider two alternatives for Taylor: adopting a questioning attitude towards the search engine, and not use the search engine in online information gathering at all.⁸⁰ I hope to show that, even if we grant that Taylor was epistemically gullible for trusting the search engine, the objection fails to undermine the extended virtue argument.

First, let us consider a questioning attitude as an alternative for Taylor's trust in the search engine.⁸¹ Following this alternative, Taylor is no longer disposed to immediately accept the search results as the most epistemically relevant search results about the COVID-19 vaccine. Taylor takes a questioning attitude towards the search engine and its search results. The feasibility of this alternative depends on how extensively Taylor questions the search engine.

One highly cumbersome way for Taylor to enact her questioning attitude is to consistently try to assess the trustworthiness of the search engine. For example, she might actively scrutinize the search results for epistemically irrelevant search results. The significant drawback of this way of questioning is Taylor would spend a lot of practical and cognitive resources on assessing the trustworthiness of the search engine. This runs contrary to the part of the motivation of using online epistemic tools in the first place. Taylor uses search engines in order to offload cognitive and practical resources. In this way, she could spend her cognitive and practical resources more wisely achieve her epistemic goals.

A more plausible way for Taylor to question is to not continually assess the search engine's trustworthiness. Instead, she continues to use the search engine, always keeping in mind that

⁸⁰ "Questioning attitude" refers to what Nguyen calls an "actively questioning attitude." Nguyen, "Trust," 25. As Nguyen writes here, this attitude is what he takes distrust to be.

⁸¹ Cf. Nguyen, "Trust," 25. See previous footnote.

the search engine results she is considering might not be the most epistemically relevant. When she considers the search result pages, she is more vigilant, trying to look for signs of epistemic relevance, but focusing on assessing the experts she has in her practicable assessment subset.

On the one hand, this approach seems much more feasible than the first way of questioning. Taylor spends less cognitive and practical resources on assessing the trustworthiness of the search engine. Moreover, the vigilance might contribute to directing trust well insofar as she might possibly be less reluctant to direct her trust towards bad experts. A vigilant mode of engagement might make Taylor more readily identify cues of untrustworthiness from purported experts, thereby preventing her from adopting an unquestioning attitude towards them.

On the other hand, Taylor still faces the problem of finding a good expert about the COVID-19 vaccine. One motivation for using search engines for Taylor was to aid her to find good experts about the COVID-19 vaccine. The idea was to use epistemically relevant search results as a guide towards finding good experts. If she cannot trust her search engine to do this, she is left with the considerable task she started with: to find good experts to trust as a novice. Though challenging, it is not necessarily an impossible task. As we will see below, Taylor could use online encyclopedias to assess the credentials of the experts she finds in her search results.⁸² Still, not having a search engine to reliably provide the most epistemically relevant results means Taylor risks spending a lot of practical and cognitive resources on experts who turn out to be bad.

Although the questioning attitude alternative is feasible, it does not show the epistemic individualist conception of the virtue of well-directed trust as the most plausible one. The line of argumentation still suggests search engines via its ordering of search results exert a considerable influence on whether Taylor winds up trusting a good expert. So, this alternative to trust has not successfully shown the individualist conception to be the most plausible conception.

However, the problem might be that Taylor uses the potentially biased search engine at all. In fact, insofar as online epistemic tool in general might be biased, one could suggest Taylor should not use them at all. This is the second alternative to trusting search engines we will consider. As an alternative to trusting online epistemic tools, Taylor should focus on building

⁸² For a case of how laypersons can use Wikipedia to assess experts, see Anderson, "Democracy," 149-53.

an epistemic network of trustworthy sources.⁸³ For example, Taylor might create a folder of trustworthy webpages: news websites, online encyclopedias, and so on.⁸⁴ So, when, she goes to these websites trying to find an expert, she consults her epistemic networks for good experts to trust. For example, she might reason that, as long as the purported is featured in the epistemic network and are not being criticized, they are a most likely a good expert. Moreover, she can verify their expertise using an online encyclopedia on the topic the expert purports to be an expert on.⁸⁵

While this can be a reasonable alternative to using search engines for achieving some epistemic goals, there are still challenges. The epistemic network might not contain the information necessary for achieve some of Taylor's epistemic goals. For example, suppose Taylor might want to learn about a very specific subject—the documentaries made by Malaysian-Taiwanese film maker Tsai Ming-liang—but her epistemic network does not contain much information. She will have to expand her online epistemic network. This is where search engines could be useful for trying to find new experts. Seeing as Taylor should not trust search engine, she consults her *offline* epistemic network—family, friends or the local librarian. But Taylor might face an analogous problem with her offline epistemic network. Suppose no one in her offline epistemic network knows about Ming-liang's filmography. Then Taylor again faces the problem of finding trustworthy experts about Tsai Ming-liang's filmography.

Most importantly, even if Taylor *could* rely entirely on her online or offline epistemic networks, it would not show the individualist conception as the most plausible conception. Her offline and online epistemic networks, I would argue, perform a similar function as the search engine did. By featuring purported experts without explicitly criticizing them, a website is signaling to Taylor that they are—at least potentially—good, trustworthy experts. If so, Taylor offloads cognitive and practical resources to her offline and online epistemic network. In effect, she is exhibiting an unquestioning attitude towards the epistemic network to provide her with potentially good experts. Similar to search engines, I would argue her epistemic network exerts a considerable influence upon how well Taylor directs her trust. By

⁸³ This line of argumentation is inspired by Nguyen's and Alfano's notion of an "epistemic network" and "testimonial networks," respectively. For an example of how Nguyen uses "epistemic network," see Nguyen, "Epistemic Bubbles," 145. For an example of how "testimonial network" apply to information acquisition via networks of news websites, see Alfano, "Virtues for agents," 8427-28.

⁸⁴ This is an adapted version of an example by Mark Alfano, mentioned in the previous footnote, where an agent acquires information from a network of news websites. Alfano, "Virtues for agents," 8427-28.

⁸⁵ Laypersons can use Wikipedia to assess experts (see footnote 82). Anderson, "Democracy," 149-53.

analogy to my argument about search engine, I take this point about online and epistemic networks to support the plausibility of the extended virtue conceptualization.

To sum up, I have argued that both alternatives to trusting search engines fail to establish the virtue of well-directed trust best conceptualized as an individualist virtue. Thus, I conclude the epistemic individualist objection has failed to undermine my main argument.

Conclusion

This chapter's aim was to answer the question: Is the virtue of well-directed trust most plausibly conceptualized as an individual or extended virtue? In Chapter 1, I identified accounting for the virtue of well-directed trust primarily in individualist terms as a condition of success for the epistemic individualist's reply to the epistemic structuralist critique. In section 1, I argued an individualist conception of the virtue of well-directed trust should be understood as an agent's disposition to adopt unquestioning attitudes towards agents and objects. However, in section 2, using the case of Taylor, I this disposition is not sufficient for accounting for directing trust well in online information gathering. I argued the search engine's bias was a considerable cause of whether Taylor wound up trusting a good expert, thus making the virtue of well-directed trust most plausibly an extended virtue. In section 3, I addressed the objection that Taylor was epistemically gullible for trusting a potentially biased search engine. I argued the objection failed. In conclusion, then, I maintain the virtue of well-directed trust is best conceptualized as an extended virtue.

This result implies that the assumption for the epistemic individualist response in Chapter 1 fails to hold. As I will argue in the concluding chapter, this suggests epistemic structuralism is the best account about explaining what makes the Internet a good source of epistemically relevant information. Yet, as I will argue, for pragmatic reasons, there is still room to be an epistemic individualist about ameliorating online information gathering. However, it would be interesting to see what follows from assuming epistemic structuralism also provides the best account for improving information gathering online. It is not obvious *a priori* how one should improve the Internet *qua* epistemic structure. This is the topic of the next chapter.

Chapter 3: Biased online epistemic tools: flawed tools or flawed online environments?*

In this chapter, I want to explore which structural improvements will improve online information gathering the most. This is an important question for epistemic structuralism about amelioration to answer. We will shift our focus away from agents and towards online epistemic tools and the online environment. This chapter will focus on an important challenge identified in Chapter 2: bias towards epistemic irrelevance in online epistemic tools.¹ This is an important problem because bias could prevent cognitively and practically limited agents from solving the problem of finding the most epistemically relevant information online.

Numerous authors have argued online epistemic tools are flawed, often focusing on the algorithms they employ. Shoshana Zuboff, for instance, argues Facebook's algorithms target polarizing information, producing what she calls "epistemic chaos."² The Wall Street Journal have argued changes to Facebook's algorithm increased engagement by promoting misinformation.³ To generalize, this suggests bias towards epistemic irrelevance is caused by a flaw with the online epistemic tool—be it a flawed algorithm or some other flaw. Let us call this "the flawed epistemic tool hypothesis."

Although the flawed epistemic tool hypothesis is an intuitive explanation, recent philosophical work provides an alternative hypothesis. This work suggests individuals and algorithms can, through no fault of their own, become socially biased due to structural problems in the environment they are embedded in.⁴ Similarly, I suggest the alternative hypothesis that an online epistemic tool becomes biased towards epistemic irrelevance simply because it is situated in an environment which structures information badly with respect to

* A special thanks goes to the GoodAttention lab meeting group at the University of Oslo, who provided helpful feedback on a late draft of this chapter.

¹ By focusing on bias towards epistemic irrelevance, I do not focus explicitly on social bias. For accounts on social bias in machine learning, see for example Gebu, "Race and Gender."; Johnson, "Algorithmic bias." Ordering information in a socially biased manner will surely lead to a bias towards presenting epistemically irrelevant information. (A fruitful comparison of the two types of biases could leverage Jessie Munton's work on salience structure and prejudice as misattributing salience to certain features of a demographic group. For a presentation of this view, see Munton, "Prejudice.") However, there are other ways an algorithm can be biased towards epistemic irrelevance not rooted in social bias. For example, an algorithm can provide highly entertaining (but *epistemically* irrelevant) information to ensure users spend time on a platform. The entertaining information need not be socially biased. Hence, I focus on bias towards epistemic irrelevance more broadly.

² Zuboff, "Facebook and the Surveillance Society." This is not to say Zuboff does not acknowledge the influence of structures, such as economic and legal ones. In the article, she proposes legal changes. My chapter focuses not on economic and legal structures, but on how online environments structure information with respect to epistemic relevance. Exploring the interaction between these various structures is beyond the scope of this thesis.

³ Wall Street Journal, "The Facebook Files, Part 4: The Outrage Algorithm."

⁴ For the influence social structures upon individuals' perceptual skill and prejudicial beliefs, see especially Munton, "Perceptual Skill," 147. For more on the influence of social structure upon social bias in machine learning algorithms, see especially Johnson, "Algorithmic bias," 9948, 9951.

epistemic relevance. In other words, the online epistemic tool can present users with epistemically irrelevant information without being flawed in any obvious way. Let us call this the “flawed epistemic relevance structure hypothesis.”

To concretize the dialectic, I have chosen to focus on Facebook and the Facebook News Feed. There are several reasons for this. First, Facebook is a platform where the creator has published documentation of how the algorithm in their online epistemic tool works, as well as how information is organized and labelled in its environment.⁵ This makes Facebook a good candidate for evaluating both hypotheses. Second, as we have seen, Facebook have been the subject of recent criticism by prominent scholars like Shoshana Zuboff. Finally, by focusing on Facebook, we also get to consider online epistemic tools other than search engines. This lets us assess whether issues from chapter 1 and 2 generalize to other online epistemic tools.

The question I want to address in this chapter is: Which hypothesis—the flawed epistemic tool hypothesis or the flawed epistemic relevance structure hypothesis—best explains the Facebook News Feed’s bias towards epistemic irrelevance? This is an important question for an epistemic structuralist about amelioration to answer. It specifies the type of structural improvement one ought to do to improve online information gathering in Facebook’s online environment. Insofar as Facebook’s environment is representative for other online environments, it indicates which structural changes ought to be done to improve online information gathering in those online environments.

The chapter is structured in the following way. To show how a flawed environment can produce bias in an epistemic tool, section 1 introduces an offline example which I call the “Veritas Library.” Offline environments provide us with intuitions helpful for deciding between the competing hypotheses in the case of Facebook. Note that since I analyze both offline and online epistemic tools in this chapter, I will use the general term “epistemic tool” when I do not need to distinguish between the two. Introducing the concepts epistemic relevance structure and epistemic order in section 2, I argue in section 3 that the librarian provides users with biased book suggestions because the library is an epistemically disordered

⁵ My chapter will draw upon Meta’s own writings about the Facebook News Feed. Lada, Wang, and Yan, "How Does News Feed Predict What You Want to See?."; Facebook, "How Does News Feed Work?," video; Mosseri, "News Feed Ranking." To better understand how information is structured on Facebook, I use Facebook’s Graph API. For an overview, see Meta, "Overview – Graph API."

Upon finishing writing this chapter, I discovered Google has published more detailed information about Google Search works and how information is structured. Google, "In-Depth Guide to How Google Search Works."; Google, "Organizing Information." I hypothesize it would be fruitful for further research to compare the information organization on Facebook with how Google Search’s online environment organizes information.

environment. This makes the flawed epistemic relevance structure hypothesis the most plausible one for the Veritas Library. Then, in section 4, I introduce simplified models of Facebook's information environment and the Facebook News Feed and map them to the Veritas Library example. In section 5, I make a conditional argument: Insofar as the simplified models of Facebook and the Facebook News Feed are representative of Facebook, the Facebook information environment is epistemically disordered. Hence, my thesis is that the flawed epistemic relevance structure hypothesis provides the best explanation of the Facebook News Feed's bias towards epistemic irrelevance. Consequently, an epistemic structuralist about amelioration ought to maintain that changing how Facebook organizes information and labels information for epistemic relevance is the best structural intervention for improving online information gathering on Facebook.

1: The Veritas Library

The aim of this section is to render plausible a crucial claim for the flawed epistemic relevance structure hypothesis. The claim is: By being embedded in an environment which orders and labels information badly with respect to epistemic relevance, an epistemic tool can order information in a biased manner—despite itself being optimized for epistemic relevance. If this claim is correct, this lends credibility to the cause of bias in this case being a flawed epistemic relevance structure, not a flaw with the epistemic tool. Although the focus of this thesis is on online environments, I will render the claim plausible by using a library as an example.⁶ Using an offline environment as an analogy will provide us with helpful intuitions for evaluating the two hypotheses in online environments like Facebook. Also, the analogy shows how some epistemic challenges in online environments generalize to offline environments. This section argues an expert librarian fails to provide users with epistemically relevant books due to the disordered state of a library. Later, in section 5, I argue a similar result holds for simplified models of the Facebook News Feed and the Facebook information environment. Consequently, under the assumption that my models are representative of Facebook, I will argue the flawed epistemic relevance structure the best hypothesis about the Facebook News Feed's bias.

Let us introduce another agent: Theo. Like the other agents we have encountered in this thesis, Theo has an epistemic goal of learning about the COVID-19 vaccine. He wants to make an informed decision about whether to take it. Also like the other agents, Theo is a

⁶ Thanks to Sebastian Watzl for suggesting and discussing library as an analogy for an online environment. Thanks also to Line Horgen Thorstad for a helpful discussion and suggestions to the analogy.

novice on the topic of vaccines and seeks information about how the vaccine works. Theo differs from the other agents in one respect, though: Theo does not consult the Internet. Instead, he goes to a well-respected research library called the Veritas Library.

Veritas Library is a special research library. It contains most books that have been printed. More books are added every day. Many of these books are highly valuable. Thus, the library has introduced a general policy: Books are stored in an archive, only accessible to librarians. In other words, unlike most libraries, users cannot find books themselves. Users must book a desk, request a librarian to find books for them, and read them in the library.

To uphold this policy, the library is organized in a particular manner. To start off, the library has hired highly specialized, expert librarians. For most knowledge domains, there is a librarian who is an expert in library science and well-versed in a knowledge domain. For example, a librarian will not only an expert in library science but is also specialized in medicine, say. So, the librarians are highly qualified to assess the epistemic relevance of books for users. In addition to specialized librarians, the library has hired archive organizers. Their task is to maintain order in the archive to help librarians effectively find epistemically relevant books for users. They organize books alphabetically and by topic. They also put topic labels on bookshelves, as well as signs in the library to help guide librarians in their search. With the help of librarians and archive organizers, then, the library has—until recently—upheld the policy about storing books in the archive.

However, Veritas Library harbors a well-kept secret. The library management believes the library has too many archive organizers. To save money, the library management has recently laid off most archive organizers. This has had disastrous consequences for the library. Books are no longer organized alphabetically and by topic. Now, they are distributed randomly on shelves. Signs and bookshelf labels are no longer reliable. A bookshelf labelled “medicine,” for instance, will no longer necessarily contain epistemically relevant books about medicine.

The library management ignored the staff’s exhortations to rehire the archive organizers. Instead, to preserve the Veritas Library’s good reputation, the management has made the librarians and remaining archive organizers sign draconian non-disclosure agreements. They are prevented from informing anyone about the disordered state of the library, lest they lose their job and be sued for breach of agreement. To cope with the disorder in the archive, the staff has created several shelves for frequently requested books, labelled “frequently

requested books.” This arrangement increases the likelihood that the librarians can keep up appearances by providing users with frequently requested books.

Let us return to Theo. Unaware of the disordered state of the library, Theo books a desk and a librarian specialized in medicine. He informs the librarian about his goal of learning about the COVID-19 vaccine. The librarian goes to the archive to find epistemically relevant books for Theo. Frantically, he looks for epistemically relevant books about vaccines. On the shelves labeled “medicine,” the librarian struggles to find any epistemically relevant books. At the shelves of frequently requested books, however, the librarian finds an older medical textbook on vaccines, as well as a popular book on the measles vaccine. The librarian puts them on a shelf trolley for potentially epistemically relevant books, which they bring with them during the search. Unbeknownst to the librarian, on a shelf labelled “technology,” there is an introductory book about mRNA vaccine technology. mRNA technology has not garnered much attention at the time Theo visits the library. So, the book has not been frequently requested by users. Moreover, given the great number of books in the library, it would be very hard for the librarian to find this book.

The librarian returns to Theo with a stack of books on the shelf trolley. The stack of books, ordered with the most epistemically relevant book on top, contains the introductory book on vaccines and the book on the measles vaccine. Importantly, the librarian’s stack of books does not contain the most epistemically relevant book in the library about the COVID-19 vaccine: the introductory book about the mRNA vaccine. Moreover, the stack is biased towards frequently requested books. Yet, popularity is not necessarily linked with epistemic relevance. Even though the introductory book to the mRNA vaccine is not frequently requested, it is arguably the most epistemically relevant book. Thus, by being biased towards popular books, the librarian’s stack becomes biased towards epistemic irrelevance.

The Veritas Library suggests two things in favor of the flawed epistemic relevance hypothesis. First, it suggests the book stack’s bias towards epistemic irrelevance is caused by the *environment*, more specifically its problematic organization and labelling of information with respect to epistemic relevance. There is no obvious flaw with the epistemic tool; the librarian does nothing obviously wrong. The librarian just happens to be in a disordered environment. It suggests the librarian’s bias is better described as a manifestation of the

disordered state of the library.⁷ Second, the Veritas Library suggests reorganizing the books in the library would ameliorate the librarian's bias. With better organization and labelling, the librarian would surely have found and added the mRNA introductory book to the stack.

Later, I will argue these two suggestions also apply to simplified models of the Facebook News Feed and the Facebook information environment. But first, to bring greater conceptual rigor to the notion of the Veritas Library being a "disordered environment," the next section introduces the terms "epistemic relevance structure" and "epistemic order." In section 3, I will use these concepts to argue that the Veritas Library is an epistemically disordered environment, thereby making the flawed epistemic relevance structure hypothesis the most plausible hypothesis about the librarian's bias. Similarly, in section 5, I argue that the flawed epistemic relevance structure hypothesis is most plausible about the Facebook News Feed's bias, as well.

2: Concepts for Flawed Epistemic Relevance Structure Arguments

To help us analyze the Veritas Library and Facebook, sections 2.1 and 2.2 introduce two concepts: epistemic relevance structure and epistemic order. Roughly, "epistemic relevance structure" captures how the environment is structured for agents and epistemic tools with respect to gathering information relevant to questions. I will define it as consisting of two components: epistemic information ordering principles and epistemic relevance cues. Building upon the epistemic relevance structure concept, I define the notion of epistemic order in section 2.2. Simply put, "epistemic order" refers to the degree to which an environment's epistemic relevance structure facilitates the inquiry necessary for answering an agent's question.

2.1: Epistemic relevance structure

I define 'epistemic relevance structure' as the manner in which an environment structures information for agents and epistemic tools to obtain epistemically relevant information for questions. As I will explain, an epistemic relevance structure is relative to a set of questions and to a set of inquiries. For example, the epistemic relevance structure of the Veritas Library refers to how written information is structured with respect to questions such as "How does the COVID-19 mRNA works?," as well as an inquiry using librarians to gather written

⁷ This idea is indebted to work by Jessie Munton and Gabrielle M. Johnson. Munton claims that an individual epistemic flaws can be a manifestation of flaws in social structures. Munton, "Perceptual Skill," 147. Johnson argues machine learning algorithms can become socially biased towards accurately modelling statistical patterns from data produced by a socially unjust environment. Johnson, "Algorithmic bias," 9948, 9951.

information. In the following, we will unpack the definition of “epistemic relevance structure.” First, however, I want to acknowledge my definition’s debt to Jessie Munton’s notion of “salience structure.”⁸

By “salience structure,” Munton refers, in part, to how an environment organizes its information for salience.⁹ Relevant for our discussion, Munton has applied the salience structure concept to the Internet, arguing it can shed light on how online environments are constructed.¹⁰ While I agree the salience structure concept can shed light on the Veritas Library and Facebook, salience is not specific enough to capture the problem.¹¹ Information can be salient without being epistemically relevant. A book can be salient because it has a striking cover, for example. Certainly, part of the problem might be that the most epistemically relevant books are not salient enough to the librarian. However, in the following, I will use the concept of epistemic relevance structure to argue the problem is more aptly described as a problem with how the Veritas Library organizes and labels information with respect to epistemic relevance.

To unpack the definition of ‘epistemic relevance structure, I will explain four key terms from the definition: environment, information, question, and inquiry. Let us start with the key term “environment.”¹² This is a difficult philosophical term to define. For the purposes of this chapter, though, I understand “environment” to be the conditions surrounding agents and epistemic tools under which they conduct their inquiries and pursue the agent’s epistemic goals.¹³ Let us illustrate the definition using the Veritas Library as an example.

The Veritas Library provides several conditions under which Theo and the librarian gather epistemically relevant information. The physical layout of the library, the information in the books, the organization of these books, labels and signs—all of these factors impact whether Theo and the librarian can successfully answer Theo’s question about the COVID-19 vaccine.

⁸ Munton, "Prejudice," 1. For an extended discussion of what salience structures are, see Munton, "Prejudice," 10-13.

⁹ Munton, "Prejudice," 1, 10-13. (See previous footnote.)

¹⁰ Munton, "Prejudice," 11n9, 17. At the GOODATTENTION opening workshop, held on May 26th, 2022, at the University of Oslo, I attended presentation by Munton, where she argued the Internet could be evaluated as a salience structure. For the abstract of Munton’s presentation, see GOODATTENTION, "GOODATTENTION: Opening Workshop."

¹¹ I thank Sebastian Watzl for encouraging me to focus on epistemic relevance. My discussion has also benefitted from Andrew Lee’s comments about the primacy of relevance over salience on Munton’s presentation at the GoodAttention opening workshop, May 26th 2022 (see previous footnote).

¹² Thanks to Sebastian Watzl for encouraging me to reflect more the concept of “environment” in the definition. This paragraph is indebted to my discussion with him on this topic.

¹³ This definition is inspired by definition 1 of “environment” in the Britannica Dictionary. Britannica Dictionary, "Environment."

This is not to say conditions outside the Veritas does not affect their ability to gather epistemically relevant information. Certainly, the accessibility of the library via the city infrastructure, whether there is a mRNA vaccine expert in the city Theo can ask, and so on, will impact whether Theo finds an answer to his question. Yet, to make the scope of this chapter manageable, I will restrict my focus to conditions provided by the Veritas Library, which I will refer to as the “Veritas Library environment.”

The second concept from the definition in need of clarification is “information.”¹⁴ The Veritas Library contains many forms of information. To start off, there is written and symbolic information. This type of information is stored in books, as well as printed on labels and signs in the library. But the Veritas Library contains other forms of information, as well. When the librarian talks to Theo, there is non-written, auditory information present in the environment.

Of course, both forms of information can answer agents’ questions. However, to make the scope of this chapter manageable, I restrict my focus to written, symbolic information. More specifically, I restrict my focus to how agents and epistemic tools leverage written and symbolic information in the environment to achieve the agent’s epistemic goals. An important motivation for this focus is that my model of the Facebook News Feed will leverage written information to predict epistemic relevance of Facebook posts for users. So, henceforth, “information” refers to written and symbolic information in the environment.

Finally, I will clarify the role of “question” and “inquiry” in the epistemic relevance structure definition.¹⁵ By “question,” I refer to the epistemic questions driving an agent’s inquiry.¹⁶ For example, the epistemic question driving Theo’s inquiry is “How does the COVID-19 mRNA vaccine work?” Answering these questions constitute an agent’s epistemic goals.¹⁷ “Inquiry” refers to the process of finding the information epistemically relevant for answering an agent’s question. Inquiry, then, is the means by which an agent answers their question,

¹⁴ Thanks to Sebastian Watzl for urging me to be specific about the type of information I will focus on.

¹⁵ My emphasis upon inquiry in the definition is influenced by a talk held by Susanna Siegel on the relationship between inquiry and salience, which I attended at the GoodAttention Opening Workshop in May 2022. For the abstract of Siegel’s talk, see GOODATTENTION, "GOODATTENTION: Opening Workshop."

¹⁶ Epistemic relevance structures apply to normative questions, as well. For example: “Is it morally defensible to take the COVID-19 vaccine when the vaccine has not been widely distributed to non-Western countries?”. An environment could be structured to find epistemically relevant information about these questions. There are several reasons why I focus on epistemic questions, though. First, the topic of this thesis is primarily epistemology. Second, epistemic questions more plausibly have a single, definite answer. For example, it strikes me as more intuitive that there is a single explanation for how the COVID-19 mRNA vaccine work. By contrast, it seems less plausible that there is a single, definite answer to the defensibility of taking the vaccine under uneven global vaccine distribution. Therefore, in the chapter, “question” always refers to epistemic questions.

¹⁷ I borrow the term ‘epistemic goal’ from Richard Heersmink, who, echoing Jason Baehr, writes that agents can have “[...] knowledge, truth or understanding” as epistemic goals. Heersmink, "Virtue Epistemology," 3.

thereby achieving their epistemic goals. For example, Theo's inquiry uses the librarian to gather epistemically relevant books about the COVID-19 vaccine. Later, we will see agents can use the Facebook News Feed in their inquiry to answer questions online, as well. To be sure, inquiry also encompasses activities of the agent. Theo's reading and reasoning about the content of the books is also part of answering his questions. Yet, as the focus of this chapter is on epistemic tools and the environment, I will focus on epistemic tools' as opposed to agents' contribution to information gathering.

I want to briefly explain why I define epistemic relevance structures relative to questions and inquiries. An environment's epistemic relevance structure can be *generalized* and *specialized*, both with respect to questions and modes of inquiry. For example, a chemistry research library can contain books from a wide range of knowledge domains. However, the books will mostly be organized with respect to answering epistemic questions about chemistry. Compare this to the Veritas Library. The Veritas Library environment orders information in a generalized manner, intended for gathering epistemically relevant information for questions in many knowledge domains.¹⁸

Similarly, an environment's epistemic relevance structure can be specialized or generalized with respect to modes of inquiry. A library can order information depending on who or what conducts the inquiry. A library where users gather epistemically relevant information without the help of librarians should organize its books and shelves without requiring prior knowledge about library science. In this case, the library would be specialized for a user-centered mode of inquiry. Conversely, a library can also order information for an epistemic tool-centered mode of inquiry. In libraries where librarians gather epistemically relevant information—e.g., the Veritas Library—the library should order information assuming expertise in library science. Doing so will make information gathering more effective for librarians. An environment's ordering of information can also be generalized with respect to inquiry—e.g., a library designed for gathering of information by users and librarians alike.¹⁹ Yet, the Veritas Library is specialized for an epistemic tool-focused mode of inquiry.

¹⁸ There is a difficulty about generalizing an environment's epistemic relevance structure with respect to questions. Optimizing a library for questions related to chemistry, say, might come at the cost of organizing the library for questions about philosophy. I suspect the library's relevance structure will have to strike a balance between different questions. I do not have the space to discuss this issue further.

¹⁹ Similar to generalizing an epistemic relevance structure for questions (see previous footnote), an epistemic relevance structure faces difficulties with respect to generalizing for modes of inquiry. It might be difficult to optimize for different modes of inquiry. Optimizing the organization of books for librarians might come at the cost of optimizing for users. Once more, I suspect the epistemic relevance structure must strike a balance. I do not have the space to discuss this issue further here.

The epistemic relevance structure of the environments we will discuss in this chapter are generalized with respect to questions and specialized with respect to modes of inquiry. As mentioned, the Veritas Library orders relevant information from numerous knowledge domains. As I will show, my simplified model of the Facebook information environment contains and structures information from numerous knowledge domains, as well. So, the epistemic relevance structures of the Veritas Library and the Facebook information environment will be generalized with respect to questions. The epistemic relevance structures are specialized with respect to modes of inquiry. Users in the Veritas Library and users in the Facebook information environment gather relevant information using epistemic tools: librarians and the Facebook News Feed, respectively.

With the central concepts of the definition explained, I want to end this section by presenting the two components of an epistemic relevance structure. The two components are: epistemic information ordering principles and epistemic relevance cues.²⁰ “Epistemic information ordering principles” refer to the principles by which the information in an environment is ordered, relative to a set of questions and modes of inquiry. For example, libraries often order books using an epistemic ordering principle of separating fiction from non-fiction. Then, using more fine-grained principles, libraries order fiction books by genre and non-fiction books by topic. Later, I will present a simplified model of Facebook where information is ordered according to principles of storing it in a social network of users, Facebook posts, and comments. So, online environments use epistemic information ordering principles, as well.

Epistemic relevance cues constitute the second component of epistemic relevance structures. An “epistemic relevance cue” is a piece of metainformation about the information stored at a location in the environment. It is a cue for epistemic relevance because it indicates the epistemic relevance of the information stored at that part of the environment. Suppose a librarian is looking for epistemically relevant books about the Vietnam war in an ordered library. They start by finding the sign for “non-fiction.” Then, the librarian looks around the non-fiction section and sees a sign that says “history.” Inspecting the shelves, the librarian finds a shelf labelled “modern American history, post WWII”. When the librarian finds a potentially epistemically relevant book, they scan the back of the book, table of contents and the index to ascertain whether the book is epistemically relevant for the Vietnam war. At every step, the librarian relies on metainformation—signs and labels in the library, table of

²⁰ Thanks to Sebastian Watzl for suggesting the notion of “relevance cue” and for discussions on the topic.

content and indices in books—to find epistemically relevant information in the environment. This applies to online environments, as well. Later, my simplified model of the Facebook News Feed will argue epistemic relevance cues play a crucial role in how well The Facebook News Feed assesses the relevance of Facebook posts.

Summing up, I have introduced epistemic relevance structures and their two components: epistemic information ordering principles and epistemic relevance cues. In the next section, I introduce the term “epistemic order,” which is the second concept we need to introduce before analyzing the Veritas Library and Facebook.

2.2: Epistemic order

I define “epistemic order” as the extent to which an environment’s epistemic relevance structure facilitates the inquiry necessary for answering an agent’s question.²¹ To illustrate the concept of epistemic order, suppose Margaret, an accountant, wants to find out how much the Veritas Library had to pay in taxes for 1960. As this was before computers were widespread in accounting firms, she must consult the epistemically relevant documents in the company filing cabinets. The epistemic order of this environment denotes how easy the ordering of financial documents in the filing cabinet makes it for Margaret to find the financial documents for 1960. In other words, it designates the ease with which Margaret can find relevant information to achieve her epistemic goal.

Intuitively, epistemic order admits of degrees. An environment’s epistemic relevance structure can make it easier or harder to find the information needed to answer an agent’s question. However, I want to introduce a distinction between what I will call “epistemically well-ordered environments” and “epistemically disordered environments.” An “epistemically well-ordered environment” is an environment with a very high degree of epistemic order.²² Let us use Margaret to illustrate. Suppose the filing cabinet has ordered financial documents in folders alphabetically by company. Each company folder has a label with the company’s name. For each company folder, documents are ordered chronologically by year and month. Margaret easily finds the relevant documents for 1960 in the company’s folder. This environment is epistemically well-ordered because its epistemic relevance structure has epistemic information ordering principles—ordering documents alphabetically and

²¹ My concept is partially the result of critical engagement with Shoshana Zuboff’s notion of “epistemic chaos.” Zuboff, “Facebook and the Surveillance Society.” Below, I outline how my concept differs from Zuboff’s.

²² One could raise the question about the threshold needed to make an environment epistemically well-ordered. As the precise threshold is not of importance to my argument, I do not determine a precise threshold here.

chronologically—and epistemic relevance cues—labels with company name—conducive for finding relevant information. This makes it easy for Margaret to find the financial documents.

Conversely, an “epistemically disordered environment” is an environment with a very low degree of epistemic order.²³ Suppose instead that the filing cabinets have few, if any, folders. Instead, the financial documents have been placed in stacks. But the stacks are neither sorted by company, nor by year or month. This makes finding the epistemically relevant financial documents a considerable challenge to Margaret. The manner in which the documents are ordered and labeled makes it severely difficult for her to find the relevant financial documents. The flawed epistemic information ordering principles and epistemic relevance cues produce a low degree of epistemic order.

Epistemic order is related to, but importantly different, from Shoshana Zuboff’s concept of “epistemic chaos.”²⁴ First off, to my understanding, “epistemic chaos” is couched in economic terms as a part of the development of “surveillance capitalism.”²⁵ My notion of epistemically disordered environment is not defined in economic terms. For another, Zuboff’s concept of epistemic chaos focuses on disinformation and misinformation.²⁶ By contrast, my concept focus on epistemic relevance. An environment can contain nothing but true information, yet still be epistemically disordered.²⁷ The financial documents in the filing cabinet might contain perfectly accurate information, but the filing cabinet can still be epistemically disordered. Finally, Zuboff thinks of epistemic chaos as caused by algorithms distributing disinformation for profit.²⁸ As I use the term, epistemically disordered environments are caused by the relevance structure’s flawed epistemic information ordering principles and epistemic relevance cues. No economic or malicious motives need to be present. For example, the disordered file cabinet in Margaret’s office can be the result of negligence of her co-workers. Epistemic order, then, is different from Zuboff’s concept in important ways.

Before proceeding, I want to briefly flag one important conceptual difficulty with the terminology I have introduced.

²³ Similar to the level of epistemic order necessary to make an environment epistemically well-ordered (see previous footnote), one could raise a similar question about the threshold necessary to make an environment epistemically disordered. Once more, a precise threshold is not crucial to my argument, so I do not provide one.

²⁴ Zuboff, “Facebook and the Surveillance Society.”

²⁵ Zuboff, “Facebook and the Surveillance Society.”

²⁶ Zuboff, “Facebook and the Surveillance Society.”

²⁷ Thanks to Sebastian Watzl for making this point to me.

²⁸ Zuboff, “Facebook and the Surveillance Society.”

2.3: Conceptual difficulties with epistemic relevance structure and epistemic order

Are both components of the relevance structure *necessary* to make an environment qualify as epistemically well-ordered or epistemically disordered? This is an important conceptual difficulty I want to flag. For space considerations, I can only briefly discuss the problem. I provide two considerations—the environment’s information-richness and capacities of the epistemic tool used—future work could use to shed more light on the difficulty.

First, I suggest information-richness affects whether we intuitively classify an environment as epistemically well-ordered. Consider the following example, which could pose a problem for my definition of epistemically well-ordered environments. Suppose an agent has a question about geometry. He is embedded in a room consisting only the five axioms of Euclidean geometry, each written on individual pieces of paper.²⁹ The papers are distributed at random, without any signs or labels in the environment to guide him. According to my definition, this environment is epistemically disordered. There are no epistemic information ordering principles and no epistemic relevance cues. Yet, the agent could gather information about the five axioms and, by deduction, answer most questions about Euclidean geometry. So, it seems to be epistemically well-ordered insofar as it facilitates the agent in answering their questions.

Information-poor environments, then, might show that the components of my definition of epistemic relevance structure are not necessary for making an environment epistemically well-ordered. However, I suspect this result will not generalize to information-rich environments. Without information-ordering principles or relevance cues, I suspect it will become difficult to find epistemically relevant information in information-rich environments like the Veritas Library or the Facebook information environment. So, even if information-poor environments are counterexamples to my definition, the two components of my definition are important for the information-rich environments we consider in this chapter.

Second, the capacities of the epistemic tool might also affect whether an environment counts as epistemically well-ordered. Consider the following example. Suppose we made changes to the epistemic relevance cues in the Veritas Library.³⁰ The library replaces human librarians with librarian drones. Every book is now given a QR-code label on the shelf on which it is placed. So, while there are no epistemic information ordering principles, the drone can scan the QR code to find the book. While this environment will not help a human librarian, it could

²⁹ Artmann, "Euclidean geometry." As this entry notes, there is some controversy regarding how these axioms ought to be formulated. To avoid needless complexity, I disregard this controversy.

³⁰ Thanks to Sebastian Watzl for suggesting a structurally similar example.

help librarian drones find epistemically relevant information. Thus, it seems that the modified Veritas Library is epistemically well-ordered *relative to the drone librarian's capacities*.

The modified Veritas Library example suggests that epistemic information ordering principles might not be necessary for making an environment epistemically well-ordered. Future work on epistemically well-ordered environments should focus on the influence of the epistemic tools' capacities. Nevertheless, in the following, I will argue both components of my definition create problems for the librarian and the Facebook News Feed.

3: Epistemically Disordered Environment as Structural Cause of Librarian's Bias

With new concepts in place, I will argue the flawed epistemic relevance structure hypothesis is the most plausible hypothesis about the cause of the librarian's bias in the Veritas Library. I will establish that the Veritas Library environment's epistemic relevance structure has such a low degree of epistemic order that it qualifies as an epistemically disordered environment. To this end, I will argue both components of the Veritas Library environment's epistemic relevance structure are seriously flawed.

Starting with epistemic relevance cues, the Veritas Library's epistemic relevance cues are not conducive to finding epistemically relevant books. The librarian cannot be confident the signs in the archive will lead to a bookshelf containing books about medicine, let alone the COVID-19 vaccine. Moreover, even if a bookshelf is labelled "medicine," it will not necessarily have epistemically relevant books about medicine on it. Flawed epistemic relevance cues partly explain why the librarian struggled to find epistemically relevant books for Theo. Hence, the epistemic relevance cue component of the Veritas Library's epistemic relevance structure is seriously flawed.

The epistemic information ordering principles component is seriously flawed, as well. In most of the library, there is no epistemic information ordering principle at all. Books are distributed at random on shelves. Yet, recall that there is a pocket of order in this anarchic environment: the shelves of popular books. Put in the terminology I have introduced, books on these shelves are ordered according to the epistemic information ordering principle of request frequency.

However, the principle of request frequency is problematic for the Veritas Library for two reasons. First and foremost, popularity does not necessarily imply high degree of epistemic relevance. The example of Theo is a case in point. The less epistemically relevant book on the measles vaccine was vastly more popular than the more epistemically relevant introductory book on the mRNA vaccine. That said, sometimes the most epistemically relevant book might

happen to be the most popular one. However, the example of Theo shows this to not always be the case.

A second problem for the principle of request frequency is the risk of creating feedback loops of epistemic irrelevance. When librarians only pick their book suggestions based on popularity, frequently requested books will be frequently suggested to new users. This cements the place of already frequently requested books on the shelf of popular books. If these books are epistemically irrelevant, this risks cementing epistemically irrelevant books on the shelf of popular books. Hence for the two reasons I have provided, the principle of request frequency is not a good epistemic information ordering principle. There is a substantial risk the principle leads the librarians' suggestions to be biased towards popularity, not epistemic relevance. So, the epistemic information ordering principle component of the Veritas Library is flawed.

As both components of the Veritas Library's epistemic relevance structure are seriously flawed, the Veritas Library environment has very low degree of epistemic order. Hence, it is an epistemically disordered environment. Moreover, there are no obvious flaws with the librarian. Had the librarian been placed in a well-ordered library, there are good reasons to think they would use their expertise to find and suggest the most epistemically relevant books to users like Theo. Thus, the flawed epistemic relevance structure hypothesis is the most plausible hypothesis about what caused the bias in the librarian's information gathering.

Consequently, this suggests the best intervention to improve information gathering in the Veritas Library is to restructure the Veritas Library environment's epistemic relevance structure. There are no obvious interventions on the librarian that will make them better. The librarian's biased information gathering seems to be a manifestation of the Veritas Library's flawed epistemic relevance structure.³¹

In the following, I will use the Veritas Library as an analogy to argue the flawed epistemic relevance structure hypothesis is also the most plausible hypothesis about the Facebook News Feed's bias. To this end, section 4 provides a simplified model of Facebook's information environment and the Facebook News Feed. In section 5, I present a conditional argument that, insofar as my models are representative of Facebook, the flawed epistemic relevance hypothesis is the best hypothesis about the Facebook News Feed's bias.

³¹ Cf. Munton, "Perceptual Skill," 147. In this paper, Munton argues well-functioning epistemic agents' epistemic flaws being a manifestation of flaws of environments with an unjust social structure.

4: From the Veritas Library to Simplified Models of Facebook

This section aims to provide simplified models of Facebook. In brief, I will understand Facebook as a social media website where users, embedded in social networks, share information with each other via Facebook posts.³² On my understanding, a user's social network consists of other users the user has "friended" or decided to "follow." The Facebook News Feed presents a user the Facebook posts from their social network by ordering posts according to relevance. Just like the librarian provided Theo with a stack with relevant books, tailored to his question, so the Facebook News Feed presents a user with a personalized ordering of relevant Facebook posts from their social network. I will expand on this brief introduction by presenting simplified models of the Facebook information environment in section 4.1, and the Facebook News Feed in section 4.2.

My simplified models will be based on publicly available documents published by Meta.³³ However, this method has limitations. First, Facebook is highly complex. So, I will have to restrict my scope. My model does not include Facebook groups, only individual users.

Another complexity is that Facebook allows users to share a wide variety of information in their Facebook posts. Users can share videos, pictures, and more. My analysis restricts the Facebook post content to written information and hyperlinks to other websites. Second, Meta's documents are often unspecific, and sometimes likely skewed towards business interests. This risks making the documents unrepresentative of how Facebook actually works.

To address both limitations, I construct simplified models with simplifying and clarifying assumptions. These are inspired by the content of the Meta documents. Yet, I do not claim they accurately represent how Facebook actually works. Nevertheless, even if the models are wrong, they can provide valuable insight. My models will contend that Facebook primarily structures information with respect to maximizing social engagement. Even if this turns out to be false for Facebook, the results still hold for other environments structured in this manner. The goal for the rest of this chapter is to provide a conditional argument: *Insofar* as these simplified models hold, I will argue the flawed epistemic relevance structure hypothesis is the most plausible one about the Facebook News Feed's bias.

³² My inspiration largely echoes definition Murthy's of social media websites. Murthy, "Theorizing Twitter," 1061. In this text, Murthy has a helpful discussion about the distinction between social media and social networks. Murthy, "Theorizing Twitter," 1061-62.

³³ For my model of the Facebook News Feed, I will primarily rely on Meta's resources. Lada, Wang, and Yan, "How Does News Feed Predict What You Want to See?"; Facebook, "How Does News Feed Work?," video; Mosseri, "News Feed Ranking." For my model of Facebook News Feed, I will primarily rely on documents from Meta's Graph API. For an overview, see Meta, "Overview – Graph API."

To concretize the simplified models and the argumentation, I will use the case of Jenny. Jenny has an epistemic goal to answer the question: What is the latest news about the COVID-19 vaccine? Jenny, like Theo, is a novice about vaccines. Therefore, Jenny conducts an inquiry by using the Facebook News Feed, an online epistemic tool, to obtain epistemically relevant information about her question. Upon logging in to Facebook, Jenny is presented a feed of ordered Facebook posts. At the top, there is a Facebook post by her relative Patrick. Patrick's post claims the COVID-19 vaccine is part of a plan to depopulate the earth, with a hyperlink to support his claim. Suspecting the point contains disinformation, Jenny and Patrick starts a heated discussion in the Facebook post's comment section. However, unbeknownst to her, Tedros Adhanom Ghebreyesus, the Director-General of the WHO, has recently published the most epistemically relevant Facebook post about the COVID-19 vaccine for her. The post is not included in her News Feed.

4.1: Simplified model of the Facebook information environment and its epistemic relevance structure

What is the Facebook environment? According to earlier definition of "environment," the answer should be the conditions surrounding Jenny and the Facebook News Feed under which they conduct the enquiry to achieve Jenny's epistemic goal. However, there is an important complication here. The precise relationship between physical environments like the Veritas Library and virtual environments like Facebook is not obvious. A commonality between the Veritas Library and the Facebook environment is they both contain written information. So, to put aside the thorny issue of the relationship between physical and virtual environments, I restrict my scope to the written information the Facebook environment contains, as well as how that information is structured. Henceforth, I will understand the Facebook as an information environment, which I will call the "Facebook information environment."

What is the Facebook information environment? According to Meta, the information on Facebook can be represented as a "social graph."³⁴ I will understand "social graph" as a network where information is stored in nodes connected by vertices.³⁵ In my simplified model of the Facebook information environment, the nodes consist of users, Facebook posts, and

³⁴ Meta, "Overview – Graph API."

³⁵ In the documents I've consulted, Meta does not provide an explicit definition of the social graph. They merely give the presentation it consists of nodes, edges and node fields. Cf. Meta, "Overview – Graph API."

comments.³⁶ The vertices represent relationships between nodes.³⁷ For example, when Jenny posts a comment on Patrick's post, this is represented in the social graph as a vertex between the comment node corresponding to Jenny's comment and the post node corresponding to Patrick's post.³⁸ Jenny's and Patrick's Facebook user profiles are represented as user nodes.³⁹

I want to focus particularly on one node type, the post node, representing a user's Facebook post.⁴⁰ As I will show, the post node is analogous to books and shelves in the Veritas Library archive. The organization of post nodes into a graph, I will argue, constitute the Facebook information environment's epistemic relevance structure. Also, the comment node stores similar kind of data, with some variations.⁴¹ Understanding the post node gives us a good understanding of the comment node, as well.

Information is stored in the post node in what I will call "post node fields."⁴² My simplified model distinguishes between two types of information stored in post node fields: post content and metainformation.⁴³ The post content is the information contained in the text of the post, as well as to hyperlinks to websites.⁴⁴ For example, the text of Patrick's post is: "As this shows, vaccination is just a part of a larger plan to get rid of us." And the hyperlink would be a link to the news article Patrick adds in support of the depopulation conspiracy.

The second type of information stored in the post node is metainformation.⁴⁵ My simplified model distinguishes between three types of metainformation in post node fields: metainformation about technical matters, about epistemic relevance, and about social engagement. Technical metainformation refers to metainformation Facebook uses for technical purposes, such as presentation to other users. For example, the post stores

³⁶ The actual social graph in the Graph API contains vastly more nodes than I use here. For an entire overview, see Meta, "Graph API Reference." For information about the post node, the primary focus here, see Meta, "Graph API Reference v14.0: Post." See also Meta, "Comment - Graph API.;" Meta, "Graph API User."

³⁷ Note that Meta calls vertices "edges" in the Graph API. Meta, "Overview – Graph API."

³⁸ Comments are listed under the post node's "edges." Meta, "Graph API Reference v14.0: Post."

³⁹ Cf. Meta, "Graph API User."

⁴⁰ Meta, "Graph API Reference v14.0: Post." However, in a later section, I will also briefly focus more on information from the user node. For an overview of the user node, see Meta, "Graph API User."

⁴¹ For a full overview of the comment node's fields and edges, see Meta, "Comment - Graph API."

⁴² 'Field' is the Graph API terminology for data stored in a node. Meta, "Overview – Graph API." For an overview of the post node's fields and edges, see Meta, "Graph API Reference v14.0: Post."

⁴³ This distinction is mine. For an overview of the post node's fields and edges, see Meta, "Graph API Reference v14.0: Post." (The last accessed version of the Graph API (January 30, 2023) lists all fields as "Node Metadata." Cf. Meta, "Overview – Graph API." My distinction might not be representative of the current version of Facebook. An updated version of my analysis should probably treat all post node fields as Node metadata.)

⁴⁴ To my understanding, in the graph API, the post content would be stored in the "message" field and the "link" field of the post node, respectively. Cf. Meta, "Graph API Reference v14.0: Post."

⁴⁵ What I call "post content" and "metainformation" are all stored in the post node's "fields". Cf. Meta, "Graph API Reference v14.0: Post." (See footnote 43 for why my distinction might no longer hold for Facebook.)

information about whether a post should be hidden from other users.⁴⁶ As my focus is not on the technical details making the Facebook information environment work, I will not focus on technical metainformation.

The two other types of metainformation stored in the post node are more pertinent to this chapter. The first type is metainformation about epistemic relevance. There are three highly pertinent pieces of metainformation about epistemic relevance: the author of the Facebook post, the publication time, as well as other users mentioned in the post.⁴⁷ In the next section, I will explain how the Facebook News Feed leverage such metainformation in its prediction about the post's relevance for users. This type of metainformation serves as a post's epistemic relevance cue. Let us call these fields "epistemic relevance fields."

The second type of pertinent metainformation pertains to social relevance. The post node stores metainformation such as whether the post is popular among users, which users are the targets of the post, and the number of times a post has been shared.⁴⁸ This type of metainformation indicates how much engagement the post has garnered from targeted users. The level of social engagement a post has created among other users can be interpreted as a cue about social relevance for a given user. Using the epistemic relevance cue as a model, let us call this type of metainformation "social relevance cues." And we can call the field the social relevance cue is stored an "social relevance field."

The post node is a good analogy to books in the Veritas Library because the post content and metainformation have clear analogies to physical books. First, the post content roughly equals the content of a physical book. The text of the post node is akin to the main body of the text in a physical book. And hyperlinks in post nodes are roughly equivalent to references and citations in physical books. For example, the introductory book to the mRNA vaccine contains references to scientific studies.

Second, metainformation in post nodes roughly equals metainformation of the physical books and shelves in the Veritas Library. The metainformation in the post node corresponds partly to a physical book's colophon, table of contents, and its index. The librarian can glean information about the author of the introductory book on the mRNA vaccine. And they can

⁴⁶ In the Graph API, there is a field called "is_hidden". Cf. Meta, "Graph API Reference v14.0: Post."

⁴⁷ The author field is based on the post node's "from" field, publication time the "created_time" field, and "message_tags" for profiles mentioned in the post. Cf. Meta, "Graph API Reference v14.0: Post."

⁴⁸ These are based on the post node fields "is_popular," "targeting," "feed_targeting," and "shares". Cf. Meta, "Graph API Reference v14.0: Post."

get an idea of the epistemic relevance by consulting the book's table of contents and index. Also, the post's social relevance cues correspond partly to the label on the frequently requested bookshelf in the Veritas Library. The librarian can gauge a book's popularity by checking whether it is placed on the shelf of frequently requested books.

With an understanding of the Facebook information environment as a social graph, we can determine the Facebook information environment's epistemic relevance structure. The social graph has both components of an epistemic relevance structure. To start off, the social graph has epistemic information ordering principles. As we have seen, the social graph store information and metainformation in post nodes and connect them via a vertex to the author's user node. These are principles by which the social graph orders various types of information into a network. The social graph also has epistemic relevance cues. We saw the post node stores a type of metainformation which serve as epistemic relevance cues. In sum, the social graph structures and labels information so that it constitutes an epistemic relevance structure. Henceforth, for readability, I will refer to the Facebook information environment's epistemic relevance structure simply as "Facebook's epistemic relevance structure."

4.2: Simplified model of the Facebook News Feed

Let us now create a simplified model of the Facebook News Feed. I will argue the Veritas Library librarian and the Facebook News Feed share similar goals and similar information ordering process to achieve their goal. This information ordering process involves two steps: finding and ranking information.⁴⁹ The primary focus is on the Facebook News Feed's ranking process, especially the machine learning algorithm I assume is involved. My model focuses on it because this algorithm could be a source of bias towards epistemic irrelevance.

⁴⁹ This distinction is inspired Meta's writings about how the Facebook News Feed works. For the steps of the Facebook News Feed involving collecting and storing candidate posts in an "inventory" and scoring them, see Lada, Wang, and Yan, "How Does News Feed Predict What You Want to See?." It is also inspired by the "ranking" described in an article by Adam Mosseri. Mosseri, "News Feed Ranking."

The reader might wonder whether dividing the information ordering process into a finding and ranking step is artificial. One might say the finding process already involves ranking, namely an assessment of whether a piece of information is a good candidate as being relevant for a user. Thus, one might conclude that distinguishing between the finding and ranking of information is artificial.

While I agree that a form of ranking is already involved in the finding step, it is still helpful to divide them into two steps. We can imagine the librarian and the Facebook News Feed operate with a certain threshold of potential relevance in the finding step. As long as the rough estimate surpasses the threshold, the book or Facebook post is added to inventory. Yet, I find it plausible to think the librarian and Facebook News Feed disregards this initial assessment in its more fine-grained relevance assessment. Even though some ranking is involved in the finding step, it is still useful to introduce a separate ranking step for the fine-grained relevance assessment.

4.2.1: The goal of the Facebook News Feed

The librarian and the Facebook News Feed share a similar goal: to order information according to relevance. The aim of the librarian is to find and order the most epistemically relevant books for users of the library. For example, for Theo, the librarian aimed to find and produce a relevance ordering of the most relevant books about the COVID-19 vaccine. Similarly, the Facebook News Feed aims to present users with “[...] content that is relevant and meaningful for them, every time they come to Facebook.”⁵⁰ For example, it aims at presenting relevant, meaningful posts to users like Jenny.

Both achieve their goal by way of an information ordering process, where they *find* information in the environment and *rank* it based on relevance for the user. The librarian provides an ordering of *books*; the Facebook News Feed algorithm presents an ordering of *Facebook posts*. In the finding step, they both create an “inventory” of potentially relevant information for the user.⁵¹ As we saw, the librarian has a shelf trolley on which they place potentially relevant books for achieving Theo’s epistemic goal. Similarly, the Facebook News Feed creates an inventory of over a thousand potentially relevant Facebook posts for a user like Jenny.⁵² In the ranking step, the librarian gives a more thorough assessment of the relevance of the books in the inventory to Theo’s goals. Similarly, in my simplified model of the Facebook News Feed, the Facebook News Feed uses a relevance score machine learning algorithm to predict a Facebook post’s relevance for a user. Let us consider each of these steps in turn.

4.2.2: The finding step

The finding step for the librarian is straight-forward, but more difficult to determine for the Facebook News Feed. The librarian finds candidate books by going through (at least parts of) the library, looking for books potentially relevant for Theo’s epistemic goal. If a book seems *prima facie* relevant, the librarian puts it on the shelf trolley. The Facebook News Feed appears to perform an analogous step of finding relevant posts for Jenna. Meta states that the inventory for a user is partly produced by finding unseen posts from other users the former

⁵⁰ Lada, Wang, and Yan, "How Does News Feed Predict What You Want to See?."

⁵¹ I borrow the term “inventory” from Meta. Lada, Wang, and Yan, "How Does News Feed Predict What You Want to See?." Elsewhere, Meta refers to the inventory as an “inventory of stories.” Mosseri, "News Feed Ranking." However, I have not managed to find a good clarification from Meta about the difference between a Facebook post and a Facebook story. Therefore, in the following, I will consider Facebook posts only.

⁵² Cf. Lada, Wang, and Yan, "How Does News Feed Predict What You Want to See?."

user “follows” or are “friends” with.⁵³ However, I have not found a detailed description of the finding process.

To provide a more substantial model to analyze later, I will expand upon Meta’s statement and proceed on the following assumption. I assume the Facebook News Feed finds Facebook posts by traversing (at least parts of) the social graph, starting with a user node. Let us use Jenny as an example to illustrate the assumption. Before Jenny logs in, the Facebook News Feed obtains candidate posts for Jenny by traversing through Jenny’s social network. The Facebook News Feed starts at Jenny’s user node, traverse the user nodes in her social network—that is, the user nodes of her “friends” and users she “follows.” For each user in the network, the Facebook News Feed checks whether the person has published a Facebook post Jenny has not seen. If she has seen the Facebook post before, the Facebook News Feed ignores it. For instance, Jenny has already seen the latest Facebook post published by her sister, so it will not be added to the inventory. By contrast, if Jenny has not seen a Facebook post, it will be added to the inventory. For example, Jenny has not seen Patrick’s Facebook post yet. Therefore, it will be added to the inventory. In this way, I assume the Facebook News Feed finds potentially relevant Facebook posts for users like Jenny.

4.2.3: The ranking step

In brief, the ranking process consists of ranking pieces of information according to relevance. The librarian and the Facebook News Feed perform the ranking step in analogous ways. Starting with the librarian, we can imagine they go through the stack of books on the shelf tray. For each book, the librarian performs a more thorough assessment of the book’s epistemic relevance for achieving Theo’s epistemic goal. For example, knowing that Theo is a novice, the librarian will regard books containing highly technical papers about vaccines as irrelevant. After this thorough assessment process, we can imagine the librarian orders the books according to epistemic relevance, and brings them to Theo.

Similarly, my simplified model of the Facebook News Feed understands the Facebook News feed as providing a relevance score to Facebook posts in the inventory for Jenny.⁵⁴ To this

⁵³ Facebook, "How Does News Feed Work?," video, 1:25-1:33; Lada, Wang, and Yan, "How Does News Feed Predict What You Want to See?": Mosseri, "News Feed Ranking."

⁵⁴ While my model focus on the relevance scoring of posts, the Facebook News Feed ranking process has more steps. Cf. Lada, Wang, and Yan, "How Does News Feed Predict What You Want to See?." It is unclear how many steps are involved. To my understanding, there are five steps to the ranking process (excepting finding candidate posts, which I described in the finding step above). Let me give a brief summary of the steps I understand to be involved. The summary is based primarily on documentation from Meta. Facebook, "How Does News Feed Work?," video; Lada, Wang, and Yan, "How Does News Feed Predict What You Want to See?."; Mosseri, "News Feed Ranking."

end, I assume the Facebook News Feed employs what I will call a “relevance score algorithm.”⁵⁵ To show how the ranking step is performed according to my simplified model, I will first explain the problem the relevance score algorithm is meant to solve. Then, I will describe how the algorithm solves the problem.

The relevance score algorithm is meant to solve what I call the “relevance score problem.” The problem is: For any given Facebook post (represented as a post node), assign an appropriate relevance score (represented as a real number between 0 and 1).⁵⁶ The relevance score algorithm solves the problem if it outputs an appropriate relevance score, for any Facebook post it takes as input. For example, suppose Jenny finds Star Trek utterly uninteresting. The algorithm should then output a very low relevance score (a real number approaching 0) for Facebook posts pertaining to Star Trek. By contrast, Jenny finds vaccine technology highly relevant, as she must decide soon whether to take the COVID-19 vaccine or not. Thus, the algorithm should output a very high relevance score (approaching 1) to posts about vaccine technology.

The first step is the initial scoring of Facebook posts. Every Facebook post in the inventory is evaluated along numerous dimensions such as the extent to which the post resembles what user typically engages with. The second step is a rule application step. To my understanding, this step ensures Facebook posts do not violate Facebook’s Community Standards. (For an overview, see Meta, “Facebook Community Standards.”) The third step is post reduction. Here, the number of posts in the inventory is reduced. As far as I can tell, this is done to save computational resources in the fourth step, the main scoring step. In this fourth step, every post is evaluated and given a single relevance score. This is the step my simplified model focuses on. The fifth and final step is the contextualization step. This is a step that ensures a user is presented with a variety of types of Facebook posts in their News Feed.

In the final stages of writing this chapter, I discovered more Meta documents pertinent to this chapter’s question. A more comprehensive analysis than mine ought to take them into consideration. First, some documents discuss how Meta fact-checks content on Facebook. (See for example Meta, “How fact-checking works.”) Second, some documents suggest content is downgraded or outright removed, if they contain misinformation. Meta, “Misinformation.”

⁵⁵ Meta refers to this as the “ranking algorithm.” Lada, Wang, and Yan, “How Does News Feed Predict What You Want to See?.” My model of the ranking algorithm builds upon the less technical presentation of machine learning algorithms by Bergstrom and West. Bergstrom and West, *Calling Bullshit*, 182-83. I have also benefitted from Gabbrielle Johnson’s more technical presentation of machine learning algorithms. Johnson, “Algorithmic bias,” 9945-49. I have also benefitted greatly from Magnus Lie Hetland’s presentation of core concepts related to algorithms and computation. See chapters 1 and 2 of Hetland, *Python Algorithms: Mastering Basic Algorithms in the Python Language*.

⁵⁶ The content of the problem is based on Meta’s public documents about how the Facebook News Feed works. Lada, Wang, and Yan, “How Does News Feed Predict What You Want to See?.”; Mosseri, “News Feed Ranking.” My formulation of the relevance score problem is indebted to Hetland’s informal definition of problems solved by algorithms. Cf. Hetland, *Python Algorithms: Mastering Basic Algorithms in the Python Language*, 10. More technically, Hetland adds, a “problem” is defined as a mathematical relation between one set (the input) and another (the output). Hetland, *Python Algorithms: Mastering Basic Algorithms in the Python Language*, 10. Accordingly, I understand the relevance score problem to be a mathematical relation between members of the set of post nodes and the set of real numbers between 0 and 1. To my understanding, an algorithm solves a problem when it, for any member of the input set, successfully provides the associated member(s) of the output set. Cf. Hetland, *Python Algorithms: Mastering Basic Algorithms in the Python Language*, 10.

Let us now specify a relevance score algorithm that allows the Facebook News Feed to solve the relevance score problem.⁵⁷ I will assume the relevance score algorithm is a machine learning algorithm. This is a reasonable assumption because Meta writes the Facebook News Feed uses machine learning technology.⁵⁸ Roughly, a machine learning algorithm is an algorithm that predicts based on “learned” statistical associations in data called “training data.”⁵⁹

In a nutshell, I assume the relevance score algorithm predicts relevance scores of Facebook posts by “learning” statistical associations between prior Facebook post and the relevance score users have assigned to them via surveys.⁶⁰ In this way, it can predict the relevance score of new Facebook post, thereby solving the relevance score problem above. Let us determine more specifically how the algorithm can solve the relevance score problem.

For our purposes, I will distinguish between two phases of the relevance score algorithm: the learning phase and the prediction phase.⁶¹ In the learning phase, the relevance score algorithm “learns” to make accurate relevance score predictions. It achieves this by training a so-called prediction model to make gradually improve predictions based on statistical patterns in “training data.”⁶² In the prediction phase, the algorithm uses the prediction model to predict the relevance score of new Facebook posts.⁶³ Before explaining these phases in greater detail, I will specify the training data and prediction model and state my assumptions about them.

I make the simplified assumption that a training datum consist of a post node and a relevance score.⁶⁴ The assigned relevance score, I assume, has been gathered via user surveys where

⁵⁷ The algorithm below is inspired by several of Meta’s presentations of the Facebook News Feed. Lada, Wang, and Yan, “How Does News Feed Predict What You Want to See?.”; Mosseri, “News Feed Ranking.”

⁵⁸ Meta states in the publicly available documentation that the Facebook News Feed employs machine learning technology. Lada, Wang, and Yan, “How Does News Feed Predict What You Want to See?.” However, as Lada, Wang and Yan add, the Facebook News Feed’s machine learning system is multi-faceted. Lada, Wang, and Yan, “How Does News Feed Predict What You Want to See?.” To make the scope of the chapter manageable, though, I treat the Facebook News Feed as if it consisted of only one machine learning algorithm.

⁵⁹ My account of machine learning algorithms is based on accounts by Bergstrom and West and Johnson. Bergstrom and West, *Calling Bullshit*, 182-83; Johnson, “Algorithmic bias,” 9945-49.

⁶⁰ Sethuraman, Vallmitjana, and Levin, “Using Surveys to Make News Feed More Personal.” As this article shows, Meta already use user surveys to personalize the News Feed for users.

⁶¹ According to Bergstrom and West’s presentation, a machine learning system uses a designated learning algorithm to produce a prediction model. Bergstrom and West, *Calling Bullshit*, 182-83. Hence, the learning algorithm and the prediction-making algorithm will likely be two different algorithms. For simplicity, though, I have decided to describe them as two phases of an overarching relevance score machine learning algorithm.

⁶² Cf. Bergstrom and West, *Calling Bullshit*, 182-83; Johnson, “Algorithmic bias,” 9945-49.

⁶³ Cf. Bergstrom and West, *Calling Bullshit*, 182-83; Johnson, “Algorithmic bias,” 9945-49.

⁶⁴ This is a simplified model because training data need not necessarily consist only of a post node. For example, one could add several fields to the training data so the algorithm could make better predictions. In footnote 75 below, I address an objection that my model focuses too much on the post node.

users provide a relevance score (a real number between 0 and 1) to the Facebook post corresponding to the post node. For example, a training datum can be a Facebook post containing information from Ghebreyesus, the Director-General of the WHO, with a user-supplied score of 0.8. Thus, the training data constitutes numerous posts nodes, each of which has user-supplied relevance scores.

Further, I make the simplifying assumption that the structure of the prediction model corresponds to the structure of a post node.⁶⁵ In other words, the prediction model has an entry for all the fields of the post node such as the popularity field, the author field. Each entry has a weight. The weight specifies the importance the prediction model gives the entry in a relevance score prediction. If the popularity entry has a high weight and the author entry a low weight, for instance, the popularity entry will play a larger role in a relevance score prediction than the author entry.

With a better understanding of the training data and the prediction model, let us now turn to the two phases of the relevance score algorithm. The learning phase consists of using training data to set the weights of the prediction model's entries.⁶⁶ I assume the algorithm uses the training data to "learn" to make accurate relevance score predictions. More specifically, I assume the algorithm tweaks the entry weights based on statistical patterns between post node fields and relevance score. For instance, if there is a strong correlation in the training data between the popularity *field* and the user-supplied relevance score, the algorithm will set a high weight to the popularity *entry* in the prediction model. The learning phase terminates, I assume, when the weights of the prediction model match the statistical associations between post node fields and the user-supplied relevance score found in the training data.⁶⁷

This takes us to the prediction phase, the second phase of the relevance score algorithm. In this phase, the relevance score algorithm uses the newly trained prediction model to make relevance score predictions about Facebook posts it has never seen before. For example, by

⁶⁵ This is a simplifying assumption because a prediction model need not necessarily have the same structure as the post node field. For one, I imagine the prediction model could make new fields in the prediction models by combining them in fruitful ways (e.g., performing mathematical operations on multiple post node fields and storing them in a new field). Cf. Google, "Feature Crosses." Alternatively, a prediction model might implement a more sophisticated statistical model than simple association between post node fields and user-supplied epistemic relevance scores. Yet, evaluating the implication of more sophisticated prediction models is beyond the scope of this chapter and my expertise.

⁶⁶ Cf. Bergstrom and West, *Calling Bullshit*, 182-83; Johnson, "Algorithmic bias," 9945-49.

⁶⁷ This is also a simplifying assumption. I suspect it might be hard for an algorithm to ascertain that it has successfully matched the *actual* statistical association in the training data. For the sake of simplicity and scope, though, I will disregard this technical problem in my model of the Facebook News Feed.

taking the fields of Patrick's post as input to the prediction model, the relevance score algorithm produces a relevance score of Patrick's post for Jenny. In this way, I assume the Facebook News Feed produces an inventory of Facebook posts for Jenny, ranked according to relevance.

The relevance score algorithm will continually move back and forth between the learning and the prediction phase. I assume the algorithm is regularly retrained on new training data. Had the algorithm only been trained once, it would risk making less accurate relevance score predictions over time.⁶⁸ While Facebook posts about the COVID-19 vaccine is relevant when Jenny logs onto Facebook now, for instance, it might be less relevant one year from now when she has already gotten her vaccine. So, to ensure the algorithm provides accurate relevance score predictions across time, the algorithm is regularly retrained.

4.2.4: Summary of the simplified model of the Facebook News Feed

Let us take stock of my simplified model of the Facebook News Feed, using Jenny as an example. The Facebook News Feed starts by *finding* posts for Jenny and place them in the inventory. It places Patrick's post, among others, into the inventory. Then, the Facebook News Feed proceeds to the *ranking* step. It provides a relevance score for Facebook posts in inventory, such as Patrick's. The Facebook News Feed orders posts in descending order of relevance score. Finally, this order is presented to Jenny when she logs into Facebook.

5: Facebook: Also an Epistemically Disordered Environment

Having created simplified models of the Facebook information environment and the Facebook News Feed, I now want to make a conditional argument. Insofar as my simplified models are representative of Facebook, Facebook is an epistemically disordered environment. By drawing comparisons to the Veritas Library, sections 5.1 and 5.2 argue both components of Facebook's epistemic relevance structure are flawed, leading to two structural causes of bias, summarized in section 5.3.⁶⁹ This suggests the flawed epistemic relevance structure the most plausible hypothesis about the Facebook News Feed's bias.

⁶⁸ This is inspired by a similar point made by Bergstrom and West. Bergstrom and West, *Calling Bullshit*, 192. They write that an algorithm aimed at classifying fake news stories would have to train on new training data to make accurate predictions across time.

⁶⁹ My idea is indebted to a similar idea by Gabrielle M. Johnson. Johnson argues that a machine learning algorithm can become socially biased simply by accurately learning statistical associations in data produced by an unjust society. Johnson, "Algorithmic bias," 9948, 9951. As I interpret her, the environment's social structure causes the machine learning algorithm's social bias.

5.1: Inadequate epistemic information ordering principles

In this section, I argue the epistemic information ordering principles in Facebook's epistemic relevance structure are flawed. We saw the social graph structures information by placing information in user and post nodes. User nodes are connected through friendship and followership. This suggests the Facebook information environment structures its information primarily for social purposes. The Facebook information environment structures information using epistemic information ordering principles conducive for socializing.

However, structuring information for socializing can lead to the Facebook News Feed not finding the most epistemically relevant Facebook post for Jenny. Recall that the most epistemically relevant Facebook post in the social graph for Jenny is Tedros Adhanom Ghebreyesus' post about the COVID-19 vaccine. We remember the Facebook News Feed finds candidate posts by traversing parts of the social graph. The Facebook News Feed will only add a Facebook post to the inventory if Jenny is friends with or follow the author of the Facebook post.

For the most epistemically relevant Facebook post to become part of her inventory, Jenny must either be friends with or follow Ghebreyesus. Let us assume for the sake of argument that Jenny is not personal friends with Ghebreyesus. The only option left, then, is that she follows him. But Jenny is a novice about COVID-19. She does not know beforehand which experts she should follow. So, it is unlikely she is a follower of Ghebreyesus. Consequently, the Facebook News Feed will not add the most epistemically relevant post for Jenny to the inventory.

Moreover, if the users in Jenny's local social network tend to post epistemically irrelevant posts, this could bias the Facebook News Feed algorithm towards epistemic irrelevance.⁷⁰ Let us assume the users Jenny follows and her Facebook friends tend to publish epistemically irrelevant Facebook posts. When the Facebook News Feed traverses the part of the social graph, then, it will find epistemically irrelevant Facebook posts. In other words, the Facebook News Feed will have a sample of Facebook posts biased towards epistemic irrelevance. What

⁷⁰ Essentially, this paragraph describes what C. Thi Nguyen calls an "epistemic bubble." For a brief discussion of the term, see Nguyen, "Epistemic Bubbles," 143-45. An 'epistemic bubble' is "[...] a social epistemic structure which has inadequate coverage through a process of exclusion by omission." Nguyen, "Epistemic Bubbles," 143; Nguyen's italics removed. In the Facebook environment, the epistemic bubble is not generated by the Facebook News Feed, but by the social principles with which the Facebook information environment orders information.

might appear as flaw with the Facebook News Feed is actually a structural problem: the posts generated by Jenny's social network being biased towards epistemic irrelevance.

One might protest that the problem is not the epistemic information ordering principles. Rather, the problem is that the Facebook News Feed traverses the social graph in a flawed manner. The protestor insists the Facebook News Feed should find posts outside Jenny's friends and users she follows. With a more adequate traversal of the social graph, the protestor argues, the Facebook News Feed will find Ghebreyesus' post for Jenny. So, the protestor concludes, the flaw is with the Facebook News Feed.

While the point about going beyond Jenny's friends and users she follows is reasonable, it is not obvious to me how the Facebook News Feed can find a more constructive traversal procedure. A naïve suggestion is to have the Facebook News Feed traverse the entire social graph. Doing so would ensure the Facebook News Feed—eventually—finds Ghebreyesus' Facebook post. However, this solution is surely not tenable. For one, given the sheer size of the social graph, traversing its entirety will likely be computationally resourceful and take a long time. For another, the inventory of potentially relevant posts would become very large. Unless the Facebook News Feed has a highly effective algorithm for assessing a giant inventory, it is implausible it can provide relevance scores to so many Facebook posts.

A more promising suggestion is to make a limited traversal beyond Jenny's friends and users she follows. In principle, it is possible the Facebook News Feed could find Ghebreyesus' Facebook post, even if Jenny is not following him. However, it is not obvious which principle the Facebook News Feed would have to follow to find Ghebreyesus' Facebook post. Insofar as my simplified model of the social graph is composed of user nodes, post nodes and comment nodes, the Facebook News Feed should surely not favor one part of the social graph over another *a priori*. Thus, the more promising solution faces substantial practical challenges, as well.

We have arrived at an analogous situation as the librarian in the Veritas Library. Just like the organization of books in the Veritas library made supremely difficult to find the most epistemically relevant book, so the information structure in the social graph makes it supremely difficult for the Facebook News Feed to find the most epistemically relevant Facebook post for Jenny. I take the lack of good alternative traversal patterns as evidence that this is not a flaw with the Facebook News Feed, but that social graph structures information

primarily for social relevance. In other words, the epistemic information ordering principles of Facebook's epistemic relevance structure are flawed.

5.2: Inadequate epistemic relevance cues

Next, I want to argue the epistemic relevance cues in Facebook's epistemic relevance structure are flawed, as well. In the Veritas Library example, the lack of good epistemic relevance cues made it difficult for the librarian to *find* epistemically relevant books. In the case of Facebook, though, I want to argue the epistemic relevance cues are inadequate for helping the Facebook News Feed to *rank* Facebook posts according to epistemic relevance.

Crucially, to accurately rank Facebook posts according to epistemic relevance, the relevance score algorithm requires information highly associated with a Facebook post's epistemic relevance. The obvious candidate is to use the information stored in the post node's epistemic relevance fields. But first I want to explain why the information in the social relevance fields are not pertinent to making good epistemic relevance predictions.

Generally, information about social relevance does not tell the relevance algorithm much about its epistemic relevance. In the Veritas Library example, we saw the most popular books about vaccine technology was not the most epistemically relevant for Theo. Similarly, a popular Facebook post will not necessarily be the most epistemically relevant one for Jenny. Patrick's conspiratorial Facebook post might become popular by generating a lot of activity from users, such as commenting. The activity need not be supportive of Patrick, though: it can merely be comments disagreeing with him.⁷¹ Admittedly, a Facebook post could also become popular because of its epistemic relevance. A Facebook post about the latest breakthrough in vaccine technology could also become popular. However, as Patrick's post illustrates, popularity is not a very reliable indicator of a Facebook post's epistemic relevance. Insofar as popularity is representative of other social relevance fields, information about social relevance is not a reliable indicator for ranking Facebook posts according to epistemic relevance.

The accuracy of the Facebook News Feed's ranking, then, hinges on epistemic relevance fields. If the epistemic relevance fields fail to make the Facebook News Feed rank Facebook posts according to epistemic relevance, they are flawed as epistemic relevance cues. Recall the relevance score algorithm's prediction model merely mimics the structure of the post node. If the Facebook News Feed's ranking is biased towards epistemic irrelevance, this

⁷¹ According to the Wall Street Journal, the promotion of outrage-generating content has been a problem for Facebook. Wall Street Journal, "The Facebook Files, Part 4: The Outrage Algorithm."

suggests the cause of the bias is structural: Facebook’s epistemic relevance structure simply has inadequate epistemic relevance cues. I will now argue that all three epistemic relevance fields introduced in section 4.1—the author, publication time, and users mentioned fields—are inadequate as epistemic relevance cues.

To start off, it is not plausible the author field provides sufficient information for the Facebook News Feed to accurately rank Facebook posts according to epistemic relevance. One might be tempted to think the author field provides ample information about the epistemic relevance of a Facebook post. For example, one could suggest knowing that the author is a medical professional is surely useful information for ranking the Facebook post. However, the author field does not contain this type of information; the only information stored is an identifier of the user who published the post.⁷² As I interpret this, the author field in the post node corresponding to, say, Patrick’s Facebook post, will only contain an identifier to Patrick’s user node.

The utility of knowing the user’s identifier is debatable. On the one hand, the information might be helpful for the Facebook News Feed. For instance, during the training phase, the relevance score algorithm might identify statistical associations between some identifiers and relevance scores. Some identifiers tend to be more associated with high relevance scores, others with low relevance scores. Based on these statistical associations, the Facebook News Feed could conceivably rank new Facebook posts according to epistemic relevance.

On the other hand, even if this were true, the strategy faces a substantial challenge. Insofar as the Facebook News Feed identifies statistical associations between unique user identifiers and relevance scores, it cannot generalize from one user to another.⁷³ That is to say, having identified a statistical pattern between the identifier for Patrick and a low epistemic relevance, it cannot generalize this pattern to other users. For this to work, it seems to me the algorithm would need to identify a statistical pattern for every user. To my understanding, this would require having representative, labelled Facebook posts for all users Jenny could encounter in the training data. In this case, the author field could in principle make the Facebook News Feed accurately rank Facebook posts. In practice, though, it would surely be unfeasible

⁷² This is based on the description of the “from” field in the Graph API, on which the author field in my model is based. Cf. Meta, “Graph API Reference v14.0: Post.”

⁷³ Here I assume unique user identifiers are randomly generated strings. If they are not randomly generated, there might be patterns between similarly generated user identifiers. If a unique user identifier encodes, say, the educational level of a user, then this could be employed by the Facebook News Feed to generalize to all users with similar levels of education.

acquire all the necessary training data. Thus, it is implausible information from the author field will make the Facebook News Feed algorithm accurately rank Facebook posts according to epistemic relevance.

The publication time field, the second epistemic relevance field, is inadequate for accurately ranking a Facebook post, as well. On the one hand, information about publication time could help ranking Facebook posts somewhat. Information on a topic like the COVID-19 changes rapidly. All else being equal, a recently published Facebook post will plausibly be more relevant than an older Facebook post. Older Facebook posts have a higher probability of containing outdated information. On the other hand, information about publication time *alone* is surely not adequate for ranking a Facebook post. Suppose Patrick's conspiratorial Facebook post was posted at the same time as a Ghebreyesus' Facebook post. Information about publication time alone will not help the Facebook News Feed rank Ghebreyesus' Facebook post higher than Patrick's. So, the publication time field by itself will not help the Facebook News Feed accurately rank Facebook posts according to epistemic relevance.

The third and final epistemic relevance field, the users mentioned in the post, is also inadequate as epistemic relevance cue. On first glance, the field looks promising. Mentioning the users of good experts—e.g., experts from the WHO—might seem like a good indicator that the Facebook post contains epistemically relevant information about the COVID-19 vaccine. Conversely, mentioning the users of bad experts—e.g., conspiracy theorists—seems to be a reliable indicator that the Facebook post is epistemically irrelevant. Upon closer consideration, though, it is not as obvious users mentioned in a Facebook post is a reliable predictor of epistemic relevance.

The problem is that mentioning bad experts is not always associated with epistemic irrelevance; and mentioning good experts is not always associated with epistemic relevance. To see why, consider the following two cases. Suppose a debunker publishes a Facebook post criticizing well-known conspiracy theorists about the COVID-19 vaccine. In the Facebook post, he only mentions users who are conspiracy theorists. Yet, despite mentioning only the bad experts, the post is arguably epistemically relevant for a user like Jenny. Next, consider a user who publishes a Facebook post containing a conspiracy about the WHO. In his post, he only mentions experts from the WHO. Still, despite mentioning only good experts, the Facebook post is epistemically irrelevant because it has conspiratorial content. Hence, the epistemic relevance field of is ill-equipped to make the Facebook News Feed reliably predict the epistemic relevance of Facebook posts.

In sum, I have argued it is implausible the three epistemic relevance fields will be adequate for ranking Facebook posts according to epistemic relevance. Thus, I take this to be decisive evidence that the second component of Facebook's epistemic relevance structure is flawed.⁷⁴

5.3: Two structural causes of the Facebook News Feed's bias

Summing up, then, I've argued that both components of Facebook's epistemic relevance structure are flawed. Like the Veritas Library example, then, the Facebook information environment is an epistemically disordered environment. Also like the Veritas Library example, this suggests the Facebook News Feed's bias towards epistemic irrelevance is caused by a flawed epistemic relevance structure.

More specifically, Facebook's epistemic relevance presents two structural causes of bias towards epistemic irrelevance for the Facebook News Feed. First, organizing information in the social graph for social purposes created a structural obstacle for the Facebook News Feed to find the most epistemically relevant Facebook post for Jenny. As we saw, it also leaves the Facebook News Feed vulnerable to adding epistemically irrelevant Facebook posts to the inventory. When a user's friends and users they follow tend to publish epistemically irrelevant posts, the Facebook News Feed adds them to the inventory. This is not perfectly analogous, but still similar to the Veritas Library example. The disordered state of the library prevented the librarian from finding the most epistemically relevant book, leaving them with no better option than to choose among frequently accessed—but potentially epistemically irrelevant—books.

Consequently, as there are no obvious ways for the Facebook News Feed to traverse the social graph to find the most epistemically relevant Facebook posts, this suggests the bias has a structural cause. The social graph should be restructured according to better epistemic information ordering principles. A tentative ameliorative suggestion is to create more nodes

⁷⁴ Someone might protest that my conclusion does not follow because I have neglected parts of the post node pertinent to epistemic relevance. The protestor argues that the post content—the text of the post and its hyperlinks—provide important cues about a Facebook post's epistemic relevance. Knowing that a Facebook post's text contains conspiratorial language or links to conspiracy theory websites would help the Facebook News Feed rank the Facebook post. The protestor insists that this shows them to be epistemic relevance cues—even though they do not contain meta-information. The protestor concludes that my argument fails because post content is sufficient to make the Facebook News Feed rank Facebook posts according to epistemic relevance.

Let us assume for the sake of argument that the protestor is correct to claim post content are epistemic relevance cues. It is still implausible the post content would make the Facebook News Feed rank adequately. Using a similar argument as we did for the epistemic relevance field about mentioning users, I would argue that conspiratorial language and links are not necessarily linked to epistemic irrelevance. A debunker might invoke conspiratorial language and link to conspiracy theory websites in order to criticize them. Hence, unless the Facebook News Feed can discern the difference between the post node content of a critic and a conspiracy theorist, I maintain the protestor's objection fails.

associated with epistemic relevance. For example, one could add topic and expert nodes. For example, one could create a topic node about COVID-19 and an expert node for Ghebreyesus. Then, since Ghebreyesus is a qualified expert on COVID-19, one could connect his expert node to the COVID-19 topic node. In addition to the users Jenny follows and are friends with, the Facebook News Feed could also traverse networks of topic and expert nodes. Intuitively, this restructured social graph would increase the probability that the Facebook News Feed finds the most epistemically relevant Facebook posts for users like Jenny.

The second structural cause of bias is the information stored in the post node's epistemic relevance fields is not highly associated with the Facebook post's epistemic relevance. We saw that this leads the Facebook News Feed to make inaccurate relevance predictions of the Facebook posts in the inventory. Even if Ghebreyesus' Facebook post were in the inventory, it would not necessarily rank it accurately as highly epistemically relevant for Jenny.

Note that the epistemic relevance cues pose different challenges to the librarian and the Facebook News Feed. While the epistemic relevance cues in the Veritas Library makes it hard to *find* epistemically relevant, it poses less of a problem for the librarian's *ranking* of the epistemic relevance of a book. The librarian has more specific info about the author and the table of contents, which have no obvious analogy in the post node's fields.⁷⁵ The epistemic relevance cues in Facebook's epistemic relevance structure, might actually make the Facebook News Feed worse off than the librarian with respect to ranking.

I want to briefly sketch a structural solution to the problem of inadequate epistemic relevance cues in Facebook's epistemic relevance structure. Facebook could add more and better epistemic relevance fields to the post node. Regina Rini has suggested social media platforms

⁷⁵ One might protest the comparison with the librarian shows my simplified model of the Facebook News Feed to be too simplistic. Had the relevance score algorithm had also access to information from the user node, the protest goes, it would make more accurate predictions. According to this view, without altering Facebook's epistemic relevance structure, we can just add more epistemic relevance cues from other nodes. While I agree that access to more information could lead to more accurate predictions, there are some *a priori* reasons to think this will not solve the problem. The field from the user node most pertinent to epistemic relevance is arguably the user's education level. Cf. Meta, "Graph API User." This could help making more accurate assessments. But this would not in and of itself solve the problem. For example, a physician can exploit their status to sell quasi-medical supplements. Moreover, relying too much on education risks underrepresenting epistemically relevant posts made by people without higher education. As I argue later in section 5.3, a better solution is to add a field pertaining to a user's track record of epistemic relevance. This is likely more associated with actual epistemic relevance of a Facebook post. It could also solve the problem of users trying to exploit their education and give more accurate epistemic relevance predictions to users without higher education. These solutions ultimately hinge on empirical results. Yet, I think these *a priori* reasons support adding new epistemic relevance cues to Facebook's epistemic relevance structure as the most plausible remedy to biased rankings of Facebook posts.

ought to keep track of users' "testimonial reputation" and computer a "Reputation Score."⁷⁶ Similarly, I suggest we could add a field for the author's epistemic relevance track record. Arguably, all else being equal, a user's track record of epistemic relevance would for predicting the epistemic relevance of future posts. Patrick has typically posted conspiratorial posts. It is not likely that he will start publishing epistemically relevant posts about COVID-19 vaccine. Conversely, Ghebreyesus has a track record of posting epistemically relevant posts about the COVID-19 vaccine. He will probably not start publishing epistemically irrelevant posts about it. This suggests that adding the author's epistemic relevance track record as an epistemic relevance field would improve the relevance score algorithm's ranking of Facebook posts.⁷⁷

Insofar as these structural problems require structural solutions, this lends credence to the flawed epistemic relevance structure as the best explanation of the Facebook News Feed's bias towards epistemic irrelevance.

Conclusion

This chapter aimed to assess which hypothesis—the flawed epistemic tool hypothesis or the flawed epistemic relevance structure hypothesis—best explains the Facebook News Feed's bias towards epistemic irrelevance. Section 1 introduced the Veritas Library. Section 2 introduced the concepts epistemic relevance structure and epistemic order. Section 3 used these to argue the Veritas Library is an epistemically disordered environment. Hence, I argued, the flawed epistemic relevance structure hypothesis best explained the Veritas Library. Section 4 introduced simplified models of the Facebook information environment and the Facebook News Feed. In section 5, I argued that, insofar as my simplified models are representative of Facebook, the Facebook information environment is also an epistemically disordered environment. For similar reasons as the Veritas Library example, I conclude the Facebook News Feed's bias towards epistemic irrelevance is best accounted for by the flawed epistemic relevance structure hypothesis. Hence, an epistemic structuralist about amelioration ought to maintain that the best structural intervention to improve information gathering on Facebook is to restructure its epistemic relevance structure.

⁷⁶ Rini, "Fake News," E57-E58. Relatedly, for an account that argues an expert's track record is an important part of a novice's expert evaluation, see especially Goldman, "Experts," 106-09.

⁷⁷ The epistemic relevance track record suggestion has several problems that would need to be addressed. To start off, Meta would first need to decide how to monitor this track record in practice. Another problem is deciding how to adjudicate a user's epistemic relevance track record across multiple knowledge domains. A user might post epistemic relevant posts about the COVID-19 vaccine, but not about election fraud. Meta would have to decide how to deal with this problem (e.g., by aggregating the relevance score across domains).

Conclusions and Reflections

The overarching question for this thesis has been: Which account—epistemic individualism or epistemic structuralism—provides the best account of the factors that makes the Internet a good source of epistemically relevant information? Recall that I distinguished between the which account provides explanation of online information gathering—the explanatory question—and which account provides the best account of how to improve online information gathering—the amelioratory question. I also restricted the scope of the questions by focusing on search engines, the Facebook News Feed, and the Facebook information environment. Below, I will give what I take to be the main results from Chapters 1 through 3.

Which account provides the best explanation of successful and unsuccessful online information gathering? I hope Chapters 1 and 2 have shown the centrality of trust accounting for gathering epistemically relevant information online. I also hope the example of Taylor using her search engine has shown that, to account for directing trust well online, we cannot merely invoke individual factors like her individual dispositions. Our account must also bring in factors pertaining to the epistemic structure she is embedded in—e.g., the epistemic network she is embedded in, how biased her online epistemic tool is towards epistemic irrelevance, and how well-ordered her online environment is with respect to epistemic relevance. If my analysis of Taylor and her use of generalize to agents using other online epistemic tools, I maintain epistemic structuralism provides the best explanatory account.

There are numerous caveats about this conclusion. For one, causal influence of online epistemic tools is crucial to the epistemic structuralist argument in this thesis. Recall from Chapter 2 that I argued in favor of the extended conception of the virtue of well-directed trust. A crucial step in the argument was the claim that search engine bias exerts a considerable causal influence upon Taylor's trust direction. While I provided reasons and a preliminary argument to believe this to be true, I have not demonstrated this thoroughly. An epistemic structuralist could reply that the bias of online epistemic tools only constitutes a background condition.

Another, more general caveat is that important Chapters 1 and 2 argued against epistemic individualism *qua* virtue responsibilism—more specifically, against Heersmink's individualist virtue responsibilism. A different epistemic individualist account might fare better against the epistemic structuralist criticism in this thesis. That said, I maintain the cases of Clara, Karl, and Taylor have demonstrated the limitations of inquirers seeking information online. Insofar as trust is crucial for agents to gather information successfully online, I think there are good

reasons to be skeptical about the prospects of epistemic individualism with respect to explanation. A final caveat is that there are different types of online environments. My thesis has focused on the familiar environments such as search engine results and social media. But there are other online environments which my thesis has not addressed—e.g., Reddit and YouTube.

Which account provides the best account of how to improve online information gathering? In brief, my thesis has not reached a definitive answer to this question. On the one hand, my thesis suggest epistemic structuralism could provide a good account. Insofar as my argument for the considerable causal influence of online epistemic tools on Internet users' trust is correct, structural improvements would plausibly improve online information gathering. And as I argued in Chapter 3, making changes to an online environment's epistemic relevance structure might be an important structural improvement.

On the other hand, there could be reasons to be an epistemic individualist about amelioration. For pragmatic reasons, it might be more fruitful to focus on Internet users. As Chapter 3 argued, fixing online epistemic tools' bias can be complicated. It might require making fundamental changes to how an online environment structures its information. Depending on the complexity of the environment, structural changes might be hard—and costly—to implement. It might be better, then, to focus on helping users able to make the best out of flawed tools and online environments.

I conclude that this thesis strongly suggests epistemic structuralism provides the best *explanatory* account of the factors making the Internet a good source of epistemically relevant information. Although the thesis suggests structural factors like an online environment's epistemic relevance structure, it has not decisively epistemic structuralism as the best *ameliorative* account.

I want to end this thesis by reflecting on an assumption of this thesis. Throughout the thesis, I have implicitly assumed the Internet should be used for answering epistemic questions. However, the Internet serve many purposes, such as meeting our social and entertainment needs.¹ Optimizing the Internet epistemic relevance could involve making a worse place to meet these non-epistemic needs. As these are important human needs, this requires us to reflect on what the Internet should be, and what we should want it to be.

¹ Thanks to Sebastian Watzl for this point and for discussions on this topic.

Bibliography

- Alfano, Mark. "Virtues for Agents in Directed Social Networks." *Synthese* 199 (2021): 8423–42. <https://doi.org/10.1007/s11229-021-03169-6>.
- Alfano, Mark, and Colin Klein. "Trust in a Social and Digital World." *Social Epistemology Review and Reply Collective* 8, no. 10 (2019): 1-8. <https://wp.me/p1Bfg0-4tk>.
- Anderson, Elizabeth. "Democracy, Public Policy, and Lay Assessments of Scientific Testimony." *Episteme* 8, no. 2 (2011): 144-64. <https://doi.org/10.3366/epi.2011.0013>.
- Arguedas, Amy Ross, Craig T. Robertson, Richard Fletcher, and Rasmus Kleis Nielsen. "Echo Chambers, Filter Bubbles, and Polarisation: A Literature Review." Reuters Institute for the Study of Journalism (*website*). Oxford University. Updated January 19, 2022, 2022, accessed March 6, 2022: 2-44. https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2022-01/Echo_Chambers_Filter_Bubbles_and_Polarisation_A_Literature_Review.pdf.
- Artmann, Benno. "Euclidean Geometry." In *Encyclopedia Britannica*, last updated January 18 2023; last accessed February 1, 2023. <https://www.britannica.com/science/Euclidean-geometry>.
- Barberá, Pablo. "Social Media, Echo Chambers, and Political Polarization." In *Social Media and Democracy: The State of the Field and Prospects for Reform*, edited by Nathaniel Persily and Joshua A. Tucker, 34-55. SSRC Anxieties of Democracy. Cambridge, United Kingdom: Cambridge University Press, 2020. <https://doi.org/10.1017/9781108890960>.
- BBC. "How Algorithms and Filter Bubbles Decide What We See on Social Media - BBC Bitesize." BBC Bitesize (*website*). BBC. Last accessed February 1, 2023. <https://www.bbc.co.uk/bitesize/articles/zd9tt39>.
- Bergstrom, Carl T., and Jevin D. West. *Calling Bullshit: The Art of Scepticism in a Data-Driven World*. London: Penguin Books, 2021. ISBN: 978-0-141-98705-7
- Berlatsky, Noah. "Google Search Algorithms Are Not Impartial. They Can Be Biased, Just Like Their Designers." NBC News (*website*). NBC News. Updated February 21, 2018; last accessed February 1, 2023. <https://www.nbcnews.com/think/opinion/google-search-algorithms-are-not-impartial-they-are-biased-just-ncna849886>.
- Britannica Dictionary. "Environment Definition & Meaning | Britannica Dictionary." The Britannica Dictionary (*website*). Encyclopædia Britannica, Inc. Last accessed February 1, 2023. <https://www.britannica.com/dictionary/environment>.

- Chun, Marvin M., Julie D. Golomb, and Nicholas B. Turk-Browne. "A Taxonomy of External and Internal Attention." *Annual Review of Psychology* 62 (2011): 73-101.
<https://doi.org/10.1146/annurev.psych.093008.100427>.
- Clark, Andy, and David Chalmers. "The Extended Mind." *Analysis* 58, no. 1 (1998): 7-19.
<https://doi.org/10.1093/analys/58.1.7>.
- Croce, Michel, and Tommaso Piazza. "Consuming Fake News: Can We Do Any Better?".
Social Epistemology (2021): 1-10. <https://doi.org/10.1080/02691728.2021.1949643>.
- Daukas, Nancy. "Epistemic Trust and Social Location." *Episteme: A Journal of Social Epistemology* 3, no. 1 (2006): 109-24. <https://doi.org/10.1353/epi.0.0003>.
- . "Altogether Now: A Virtue-Theoretic Approach to Pluralism in Feminist Epistemology." In *Feminist Epistemology and Philosophy of Science: Power in Knowledge*, edited by Heidi E. Grasswick, 45-67. Dordrecht: Springer, 2011.
<https://doi.org/10.1007/978-1-4020-6835-5>.
- Dyrendal, Asbjørn, and Terje Emberland. *Hva er konspirasjonsteorier*. Oslo: Universitetsforlaget, 2019. ISBN: 978-82-15-02993-1.
- Editors of Encyclopaedia Britannica. "Algorithm." In *Encyclopedia Britannica*, updated May 6 2021; last accessed February 1, 2023.
<https://www.britannica.com/science/algorithm>.
- Facebook. "How Does News Feed Work?". *Facebook*, October 20, 2017. Video, 2:51. Last accessed February 1, 2023.
<https://www.facebook.com/facebook/videos/10156625242821729/>.
- Fricker, Miranda. *Epistemic Injustice : Power and the Ethics of Knowing*. Oxford: Oxford University Press, 2007. <https://doi.org/10.1093/acprof:oso/9780198237907.001.0001>.
- . "What's the Point of Blame? A Paradigm Based Explanation." *Noûs* 50, no. 1 (2016): 165-83. <https://doi.org/10.1111/nous.12067>.
- Gebri, Timnit. "Race and Gender." In *The Oxford Handbook of Ethics of AI*, edited by Markus D. Dubber, Frank Pasquale and Sunit Das, 253-69. Oxford Handbooks. New York: Oxford University Press, 2020.
<https://doi.org/10.1093/oxfordhb/9780190067397.013.16>.
- Goldman, Alvin I. "Experts: Which Ones Should You Trust?". *Philosophy and Phenomenological Research* 63, no. 1 (2001): 85-110. <https://doi.org/10.1111/j.1933-1592.2001.tb00093.x>.
- GOODATTENTION. "GOODATTENTION: Opening Workshop - Department of Philosophy, Classics, History of Art and Ideas." GOODATTENTION research project

- (*website*). University of Oslo. Updated May 24, 2022; last accessed February 1, 2023. <https://www.hf.uio.no/ifikk/english/research/projects/goodattention/events/Workshops%20and%20Conferences/goodattention-opening-workshop.html>.
- Google. "Feature Crosses: Crossing One-Hot Vectors | Machine Learning | Google Developers." Google Developers (*website*). Google. Updated July 18, 2022; last accessed February 1, 2023. <https://developers.google.com/machine-learning/crash-course/feature-crosses/crossing-one-hot-vectors>.
- . "In-Depth Guide to How Google Search Works | Google Search Central | Documentation | Google Developers." Google Search Central (*website*) Google. Updated May 26, 2022; last accessed February 1, 2023. <https://developers.google.com/search/docs/advanced/guidelines/how-search-works>.
- . "Organizing Information — How Google Search Works." Google Search (*website*). Google. Last accessed February 1, 2023. https://www.google.com/intl/en_us/search/howsearchworks/how-search-works/organizing-information/.
- . "Our Approach – How Google Search Works." Google Search (*website*). Google. Last accessed February 1, 2023. https://www.google.com/intl/en_us/search/howsearchworks/our-approach/.
- Heersmink, Richard. "A Virtue Epistemology of the Internet: Search Engines, Intellectual Virtues and Education." *Social Epistemology* 32, no. 1 (2018): 1-12. <https://doi.org/10.1080/02691728.2017.1383530>.
- Hetland, Magnus Lie. *Python Algorithms: Mastering Basic Algorithms in the Python Language*. 2 ed. Berkeley, CA: Apress, 2014. <https://doi.org/10.1007/978-1-4842-0055-1>.
- Johnson, Gabrielle M. "The Structure of Bias." *Mind* 129, no. 516 (2020): 1193-236. <https://doi.org/10.1093/mind/fzaa011>.
- . "Algorithmic Bias: On the Implicit Biases of Social Technology." *Synthese* 198 (2021): 9941–61. <https://doi.org/10.1007/s11229-020-02696-y>
- Kahneman, Daniel. *Thinking, Fast and Slow*. London: Penguin Books, 2012. ISBN: 978-0-141-03357-0.
- Kawall, Jason. "Other-Regarding Epistemic Virtues." *Ratio* 15, no. 3 (2002): 257-75. <https://doi.org/10.1111/1467-9329.00190>.
- Lada, Akos, Meihong Wang, and Tak Yan. "How Does News Feed Predict What You Want to See?" *Meta* (*website*). Meta. Published January 26, 2021; last accessed February 1,

2023. <https://about.fb.com/news/2021/01/how-does-news-feed-predict-what-you-want-to-see/>.
- Leonard, Nick. "Epistemological Problems of Testimony." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. Stanford, CA: The Metaphysics Research Lab, Philosophy Department, Stanford University, 1997-. Article first published April 1, 2021. <https://plato.stanford.edu/archives/sum2021/entries/testimony-episprob/>.
- Levy, Neil, and Mark Alfano. "Knowledge from Vice: Deeply Social Epistemology." *Mind* 129, no. 515 (2020): 887-915. <https://doi.org/10.1093/mind/fzz017>.
- Lorenz-Spreen, Philipp, Lisa Oswald, Stephan Lewandowsky, and Ralph Hertwig. "Digital Media and Democracy: A Systematic Review of Causal and Correlational Evidence Worldwide." Working paper submitted November 22, 2021 to *SocArXiv*, accessed March 6, 2022. <https://doi.org/10.31235/osf.io/p3z9v>.
- McLeod, Carolyn. "Trust." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. Stanford, CA: The Metaphysics Research Lab, Philosophy Department, Stanford University, 1997-. Article first published February 20, 2006; last modified September 15, 2021. <https://plato.stanford.edu/archives/fall2021/entries/trust>.
- Meta. "Comment - Graph API." Meta for developers (*website*). Meta. Last accessed February 1, 2023. <https://developers.facebook.com/docs/graph-api/reference/v14.0/comment>.
- . "Facebook Community Standards | Transparency Center." Meta Transparency Center (*website*). Meta. Last accessed February 1, 2023. <https://transparency.fb.com/policies/community-standards/>.
- . "Graph API Reference." Meta for developers (*website*). Meta. Last accessed February 1, 2023. <https://developers.facebook.com/docs/graph-api/reference>.
- . "Graph API Reference V14.0: Post." Meta for developers (*website*). Meta. Last accessed February 1, 2023. <https://developers.facebook.com/docs/graph-api/reference/post/>.
- . "Graph API User." Meta for developers (*website*). Meta. Last accessed February 1, 2023. <https://developers.facebook.com/docs/graph-api/reference/v14.0/user>.
- . "How Fact-Checking Works | Transparency Center." Meta Transparency Center (*website*). Meta. Updated October 4, 2022; last accessed February 1, 2023. <https://transparency.fb.com/en-gb/features/how-fact-checking-works/>.
- . "Misinformation | Transparency Center." Meta Transparency Center (*website*). Meta. Last accessed February 1, 2023. <https://transparency.fb.com/policies/community-standards/misinformation/>.

- . "Overview – Graph API." Meta for developers (*website*). Meta. Last accessed February 1, 2023. <https://developers.facebook.com/docs/graph-api/overview>.
- Mosseri, Adam. "News Feed Ranking in Three Minutes Flat." Meta (*website*). Meta. Published May 22, 2018; last accessed January 30, 2023. <https://about.fb.com/news/2018/05/inside-feed-news-feed-ranking/>.
- Munton, Jessie. "Perceptual Skill and Social Structure." *Philosophy and Phenomenological Research* 99, no. 1 (2019): 131-61. <https://doi.org/10.1111/phpr.12478>.
- . "Prejudice as the Misattribution of Saliency." Marc Sanders Foundation (*website*). Marc Sanders Foundation, accessed December 22, 2021: 1-19. <https://marcsandersfoundation.org/wp-content/uploads/2020/11/Prejudice-as-the-misattribution-of-saliency.pdf>.
- Murthy, Dhiraj. "Towards a Sociological Understanding of Social Media: Theorizing Twitter." *Sociology* 46, no. 6 (2012): 1059–73. <https://doi.org/10.1177/0038038511422553>.
- Nguyen, C. Thi. "Cognitive Islands and Runaway Echo Chambers: Problems for Epistemic Dependence on Experts." *Synthese* 197, no. 7. (2020): 2803-21. <https://doi.org/10.1007/s11229-018-1692-0>.
- . "Echo Chambers and Epistemic Bubbles." *Episteme* 17, no. 2. (2020): 141-61. <https://doi.org/10.1017/epi.2018.32>.
- . "Trust as an Unquestioning Attitude." Version 2. PhilArchive (*website*). The PhilPapers Foundation, uploaded November 2, 2020: 1-48. <https://philarchive.org/archive/NGUTAAv2>.
- . "The Seductions of Clarity." *Royal Institute of Philosophy Supplements* 89 (2021): 227-55. <https://doi.org/10.1017/S1358246121000035>.
- NumFOCUS, Inc. "Pandas.DataFrame.Describe — Pandas 1.5.3 Documentation." pandas (*website*). NumFOCUS, Inc. Last accessed February 1, 2023. <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.describe.html>.
- O'Neill, Onora. "Trust and Accountability in a Digital Age." *Philosophy* 95, no. 1 (2020): 3-17. <https://doi.org/10.1017/S0031819119000457>.
- Origgi, Gloria. "Is Trust an Epistemological Notion?". *Episteme* 1, no. 1 (2004): 61-72. <https://doi.org/10.3366/epi.2004.1.1.61>.
- . "Epistemic Injustice and Epistemic Trust." *Social Epistemology* 26, no. 2 (2012): 221-35. <https://doi.org/10.1080/02691728.2011.652213>.

- Pasquale, Frank. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge, Massachusetts: Harvard University Press, 2016. ISBN: 978-0-674-97084-7.
- Rini, Regina. "Fake News and Partisan Epistemology." *Kennedy Institute of Ethics journal* 27, no. 2 Supplement (2017): E-43-E-64. <https://doi.org/10.1353/ken.2017.0025>.
- Schwengerer, Lukas. "Online Intellectual Virtues and the Extended Mind." Version 2. PhilArchive (*website*). The PhilPapers Foundation, uploaded October 15, 2020: 1-21. <https://philarchive.org/archive/SCHOIVv2>.
- Sethuraman, Ramya, Jordi Vallmitjana, and Jon Levin. "Using Surveys to Make News Feed More Personal | Meta." *Meta (website)*. Meta. Updated July 2, 2020; last accessed February 1, 2023. <https://about.fb.com/news/2019/05/more-personalized-experiences/>.
- Smart, Paul R., and Robert W. Clowes. "Intellectual Virtues and Internet-Extended Knowledge." *Social Epistemology Review and Reply Collective* 10, no. 1 (2021): 7-21. <https://wp.me/p1Bfg0-5AY>.
- Sperber, Dan, Fabrice Clément, Christophe Heintz, Olivier Mascaro, Hugo Mercier, Gloria Origgi, and Deirdre Wilson. "Epistemic Vigilance." *Mind & Language* 25, no. 4 (2010): 359-93. <https://doi.org/10.1111/j.1468-0017.2010.01394.x>.
- Turri, John, Mark Alfano, and John Greco. "Virtue Epistemology." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. Stanford, CA: The Metaphysics Research Lab, Philosophy Department, Stanford University, 1997-. First published July 9, 1999; last modified October 26. <https://plato.stanford.edu/archives/win2021/entries/epistemology-virtue/>.
- Wall Street Journal. "The Facebook Files, Part 4: The Outrage Algorithm - the Journal. - WSJ Podcasts." *The Facebook Files (website)*. The Wall Street Journal. Updated September 18, 11:30 AM, 2021; last accessed February 1, 2023. <https://www.wsj.com/podcasts/the-journal/the-facebook-files-part-4-the-outrage-algorithm/e619fbb7-43b0-485b-877f-18a98ffa773f>.
- Wilson, Deirdre, and Dan Sperber. "Relevance Theory." In *The Handbook of Pragmatics*, edited by Laurence R. Horn and Gregory Ward, 607-32. Blackwell Handbook in Linguistics. Malden, MA: Blackwell Publishing, 2006. <https://doi.org/10.1002/9780470756959.ch27>.
- Woodward, James. "Causation in Biology: Stability, Specificity, and the Choice of Levels of Explanation." *Biology & Philosophy* 25, no. 3 (2010): 287-318. <https://doi.org/10.1007/s10539-010-9200-z>.

Zuboff, Shoshana. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. London: Profile Books, 2019. ISBN: 978-1-78125-685-5.

———. "Facebook and the Surveillance Society: The Other Coup." *The New York Times Opinion (website)*. The New York Times. Published January 29, 2021; last accessed February 1, 2023. <https://www.nytimes.com/2021/01/29/opinion/sunday/facebook-surveillance-society-technology.html>.

Øverby, Harald, and Henrik Dvergsdal. "internett." In *Store norske leksikon*, edited by Erik Bolstad, Foreningen Store norske leksikon. Last modified October 1, 2021; last accessed February 1, 2023. <https://snl.no/internett>.