The Design of a Critical Thinking Course as a Philosophical and Pedagogical Project

by
Lyle Crawford

Master of Arts, University of Victoria, 2004
Bachelor of Arts, University of Victoria, 2002

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

in the
Department of Philosophy
Faculty of Arts and Social Sciences

SIMON FRASER UNIVERSITY
Spring 2023

Declaration of Committee

Name:	**Lyle Crawford**
Degree:	**Doctor of Philosophy**
Title:	**The Design of a Critical Thinking Course as a Philosophical and Pedagogical Project**
Committee:	**Chair: Endre Begby** Associate Professor, Philosophy
	Nicolas Fillion Supervisor Associate Professor, Philosophy
	Evan Tiffany Committee Member Associate Professor, Philosophy
	Jennifer Wang Committee Member Associate Professor, Philosophy
	Bruno Guindon Committee Member Lecturer, Philosophy
	Tom Archibald Examiner Professor, Mathematics
	Michael Picard External Examiner Professor, Philosophy Douglas College

Abstract

Critical thinking (CT) is an important learning objective in most liberal education. In post-secondary education, CT is often taught as its own subject with a dedicated first-year course, normally in the philosophy department. I describe the design of such a course, address some major philosophical questions about CT and CT pedagogy, and present a textbook I created that is suitable for students with diverse educational and language backgrounds. The course design reflects several years of trial and error in college and university classrooms as well as a designation at Simon Fraser University as a quantitative and a science course. I present a picture of the critical thinker as someone with skills and dispositions that prioritise truth, the rational use of evidence, and clarity of thought and communication. I take seriously but reject the contentions that CT encourages combativeness and that CT is an inherently subject-specific activity that cannot be taught in a general standalone course.

I describe the course design with a set of parameters reflecting oppositions or trade-offs. The course adopts a restrained pluralism in its approach to reasoning, is focused on skills over dispositions, balances comprehension and communication skills, emphasizes reactive thinking skills but not to the exclusion of constructive ones, and employs an extremely concise writing style. I also describe the course design in terms of specific content. I show that my reconstruction of some common arguments as short extended patterns helps students to understand the language and commitments of these arguments and to criticize them in useful ways. I defend a highly pragmatic treatment of deductive reasoning. I defend the course's focus on arguments against some actual and potential objections informed by cognitive science, Paul Thagard's critique of a CT based on informal logic, and a concern that arises from Hugo Mercier and Dan Sperber's evolutionary theory of reasoning. Finally I explain how my treatment of Bayes' rule relates to my treatment of scientific reasoning, and I show that my partially original geometric teaching method captures what is best in the traditional form of Bayes' rule and in an increasingly popular but limited alternative form.

Keywords: critical thinking; arguments; Bayes' rule; informal logic

Table of Contents

Introduction

Critical thinking (CT) is among the most widely and strongly affirmed learning objectives of a liberal education. The most expansive notions of CT encompass a wide range of desirable knowledge, skills, dispositions, values, and character traits. Development of students' CT is routinely promised by courses and programs throughout higher education, as well as K-12, with various understandings of what that promise entails. In addition, most colleges and universities offer a course explicitly devoted to Critical Thinking, normally run by the Philosophy Department. To the extent that CT is acknowledged as an independent subject that can be directly studied in a generalized form – something not universally granted even within philosophy – philosophy unofficially owns CT because some of the principal tasks of CT are closely related to traditional concerns of philosophy going back to its ancient roots. These include the execution, communication, and evaluation of reasoning about what to believe and do.

This dissertation presents, describes, contextualizes, and defends a CT course offered by the Department of Philosophy at Simon Fraser University, Canada, as well as at several nearby colleges. The single largest component (Appendix C) is the document that serves as the foundation of this course, a freely distributed textbook (described as a "course manual") designed to be maximally efficient and accessible to students with a wide range of academic and language backgrounds. A few of the corresponding highly developed slide resources are presented in demonstration videos (Appendices D-F).

The approach to CT and CT pedagogy that I lay out in this project is based on several years of reflection and trial-and-error in college and university classrooms. During these years I discovered that many available CT textbooks, while offering things of value for *me* to read in contemplating and designing a CT course, were or little value to typical and even high-performing students at these institutions. I find nothing among the published CT offerings that is as useful to these students as is the course manual offered here, in its writing and visual style, and in its overall construction.

The course materials in the appendices are mostly not offered as evidence for freestanding theses about CT. Presenting and defending the course is a major task of this dissertation, which is why it is a pedagogical project as much as a philosophical one. I don't present and defend the course because it is a radical or idiosyncratic approach that defies philosophical orthodoxy, but because it is a distinctive, effective, refined, and extensively classroom-tested solution to a problem –

enhancing the CT of a particular group of learners – recognized by institutions of higher education, by philosophy departments, and by society generally. The overall goal of this project is to present a well-designed course that introduces students to some of the key tools and standards of good reasoning that they should expect to see and use in other philosophy courses, in their academic career, and in their lives generally, and to defend its merits with reference to some of the active discussions in the growing CT literature (e.g. Blair 2021). The textbook (including an online Bayes' rule widget) is a free resource available to anyone attracted to its highly concise style and persuaded of the value of a CT course in which Bayesian reasoning is prominent but not all-encompassing.

Although this dissertation is not organized as a list of direct answers to questions about the course, there *is* a set of questions about the course that I do aim to answer, in some cases fairly directly and in other cases through an exploration of a relevant issue in the scholarship around the philosophy of CT and CT education. The first question concerns the very existence of the course; the other four concern the contents and structure and role of the course.

1) Why have a standalone CT course, a course dedicated to CT as an independently learnable subject? And how does this course make the most of the limited opportunity of a single course to help students become critical thinkers? This question includes the concern that such a course is misguided or futile, as well as the question of what aspects of becoming a critical thinker a standalone course might profitably address.

2) How does this course and course manual serve first-year level students with a wide range of linguistic and educational backgrounds?

3) How does this course serve as a quantitative skills course? This question refers to the course's designation, by Simon Fraser University, as one of several basic quantitative courses students may choose from in order to fulfill their degree requirements.

4) How does this course serve as a science / social science course? Like the previous question, this refers to a designation by Simon Fraser University for its degree requirements.

5) How does this course function as a philosophy course? How does it represent philosophy to the majority of students who will not do more academic philosophy, and how does it contribute to the philosophical development or preparation of the minority who will? This

question includes several issues around the course's treatment of the fundamental philosophical device, the argument, as well as a concern, which arises sometimes for academic philosophy in general, that CT promotes socially undesirable traits such as intellectual combativeness.

This dissertation does not, as I said, take the form of a list of answers to these questions. My aim is for answers to emerge throughout a three-part plan.

Part 1 focuses mostly on the educational and social context of CT. This includes the deeper historical and philosophical background of CT and informal logic, the problem of defining CT, and some important related concerns: whether an education designed to promote CT inadvertently promotes problematic intellectual traits such as combativeness, and whether CT is even teachable in the format of a separate, dedicated course. I contend that CT is in fact teachable in this form, and I draw some lessons, applicable to the design of this course, from the debate on this issue. Part 1 addresses aspects of questions (1), (2), and (5). It looks at how the expectation developed, within the education system and to some extent within society generally, that CT development is a job for formal higher education, and a job partially but significantly achievable with a special-purpose course taught by Philosophy. I'll look at the development of CT from, in, and adjacent to the philosophical tradition of concern for reliable methods of truth-discovery and rational methods of persuasion.

Part 2 describes the design of the course in terms of oppositions or trade-offs which I'll call "parameters" that reflect both philosophical questions about the character of CT and pedagogical questions about how to respond to constraints imposed by the format of a single-term course taken by large numbers of first-year and international students. I discuss some of the course's content – the topics, tools, and activities of CT – in Part 2 to illustrate and explain the parameter "settings". Part 2 addresses aspects of questions (2) and (5). The final parameter, writing style, directly concerns the question of the course's accessibility to students of diverse backgrounds, in particular linguistic backgrounds. This reflects a concern that the course be suitable for use in an area such as the lower mainland of British Columbia with a particularly diverse student body, and also that it be flexible enough to function well at multiple institutions, offering educational value to students whose academic preparation and ambitions vary considerably. The first parameter, the unity of the approach to reasoning, concerns the more philosophical question of how strongly to commit to the "toolkit" metaphor of CT skills and whether doing so is a pragmatic decision or reflects a real deep diversity of good reasoning methods. The overall function of Part 2 is to describe how the course balances priorities that compete with one another in the context of a one-term introduction to CT.

Part 3 covers other major elements of the course content, though it is not exhaustive. The first element is the significant investment in argument as a tool for the composition and communication of reasoning. To the extent that I focus on argument analysis and informal logic, this is not merely because these are traditional for CT. I consider some actual and potential critiques of this focus that are informed by cognitive science. Although I defend a moderate focus on argument, the course is responsive to some of these concerns, for example in its attention to the most pervasive forms of cognitive bias instead of to the traditional fallacies. The final section is where I most directly discuss some of the ways in which the course earns its designations at SFU as a Quantitative and as a Science course, addressing questions (3) and (4). I explain some aspects of a selective and pragmatic treatment of deductive reasoning that avoids “logic for logic’s sake”, and I discuss my treatment of the reasoning traditionally called *abductive*. I also show how my partially original geometric method of teaching the course’s most important quantitative thinking tool, Bayes’ rule, harnesses an effective mechanism of visual learning. This method incorporates the best elements of two alternative approaches, one that is insufficiently sensitive to the unsteady quantitative skills that many students bring to a first-year Arts course, and one that is superficially simpler but unattractively limited.

Part 1. Critical Thinking and Education

Introduction

In this part of this project on the design of a critical thinking (CT) course, I'll look at some relevant history, definitional questions, and concerns about the character of CT; and I'll address a longstanding question about CT with significant implications for how CT could be taught. These are academic issues, certainly, but CT itself is of much more than academic concern. I'll begin with a few non-academic affirmations of the great importance of CT. I take these to be typical non-academic uses of the term *critical thinking*. In them we can see many of the issues with which I'll be concerned in this project.

❖

"Can Randi Weingarten Save Public Schools?" (*New York Times*, 2021 Dec 17)

"In August, *Los Angeles Magazine* profiled Cecily Myart-Cruz, head of that city's teachers union, who insisted that there was no such thing as pandemic-related learning loss. "It's OK that our babies may not have learned all their times tables," she said. "They learned resilience. They learned survival. They learned **critical-thinking** skills. They know the difference between a riot and a protest" [referring to the George Floyd protests].

❖

"How an Electric Truck Factory Became a Lightning Rod in Georgia" (*New York Times*, 2022 Mar 14). Subheadline: "The governor hailed the factory as an economic boon that would put Georgia at the vanguard of the green economy. Not everyone liked the idea."

Reader comment: James (New Mexico): "There is a desperate need for **critical thinking** classes in the Georgia School systems."

❖

"Virginia Republicans Want to Strip **Critical Thinking** from the Classroom" (*Slate*, 2022 Jan 15)

A proposed bill would require teachers to either avoid or take a "neutral", "multiple viewpoints" approach to any of several issues of concern to conservatives. The article concludes: "It's not just unflattering facts about the Founding Fathers that Republicans are trying to keep out of public schools. It’s **critical thinking** itself.”

This final sentence of the article is the sole reference to CT.

❖

“Why Sex Education is So Bad in the U.S.” (CNBC YouTube, 2021 Dec 21)

The CEO of Ascend, a US advocacy group that opposes comprehensive sex education for teenagers and promotes abstinence until marriage, describes her group’s mission: “It’s really important that we reinforce those good habits that the teens are making right now by helping them with refusal skills, self-regulation, helping them with goal-setting, putting an eye on their future — personal agency is extremely important and we’re only there to provide the medical facts and to also begin to instill some **critical thinking**.”

❖

“Don’t Just Watch: Team Behind “Don’t Look Up” Urges Climate Action” (*New York Times*, 2022 Jan 11)

The satirical film uses an imminent comet collision as an exaggerated metaphor for climate change. The director, Adam McKay, says: “If it inspires conversation, **critical thinking**, and makes people less tolerant of inaction from their leaders, then I’d say we accomplished our goal.”

❖

It can sometimes be tempting to take a cynical view of comments like these and read “critical thinking” as little more than a way for people to say “opinions/attitudes that I agree with”. To some extent, that’s probably fair. The comments may reflect a superficial grasp of the concept of CT as well as an over-confidence that others’ failure to think as the authors do is due to insufficiently critical thinking. Indeed, CT is generally said to be something that *other* people should be better at. Even so, there are some interesting ideas in these comments that set the stage for my project.

1) The term *critical thinking* has become a familiar piece of everyday language outside of academic and pedagogical contexts. It’s used with confidence that others will immediately understand what is meant.

2) Even with their full contexts, it would be difficult to derive any single or simple definition from these uses of the concept. In everyday usage, the concept is vague and its meaning shifts.

3) CT is thought to be a good thing that more people should do more of.

4) CT is taken to consist of general skills that may, and perhaps may only, be learned through direct instruction or life experiences such as those that demand adaptability or resilience.

5) CT is thought to be something that people can learn in a dedicated course on the subject.

6) CT is seen as something like a character trait, not just an activity.

7) CT is seen as directed towards the formation of true beliefs and the avoidance of false ones.

8) CT is seen as directed toward, and reflected in, good decisions and actions.

9) CT is seen as something from which society as a whole, as well as the individuals within in it, stand to benefit. Society is right to expect and encourage its members to attain a certain level of CT. As the CT skeptic Daniel Willingham puts it, "in free societies, the ability to think critically is viewed as a cornerstone of individual civic engagement and economic success" (2020: 41).

Most of these ideas are also discernible in systematic surveys of the ways CT is understood by the public, and not just in North American or English-language contexts. For example, a recent study (Penkauskienė *et al*. 2019) of attitudes toward *critical thinking* (and corresponding terms in European languages) among employers in Europe finds that views about CT, although invariably positive, are also "general and ambiguous".[1] It's seen as a capacity for getting the right answer and making the right decision, for both "personal improvement and common good" (2019: 804).

In academic contexts, the widespread judgement that not enough people are doing enough CT takes the form of fears that education systems are failing to make students into critical thinkers. In 1941, Robert Glaser, an important figure in the 20th century development of CT education whom we'll meet again in section 1.1.2, was motivated by his perception that the then-recent success of literacy education had not been matched by success in getting to people to critically evaluate the things they were newly able to read (1941: 4-5).

[1] *Critical thinking* is certainly not the only pedagogical concept that receives near-universal endorsement despite, or perhaps in part because of, the fact that people mean different things by it. For example, another is *active learning* (Sanders *et al*.: 2017).

More recently, the 1983 report of the US National Commission on Excellence in Education, *A Nation at Risk*, sounded the alarm about many ways in which students of the American K-12 education system were under-performing, in particular with higher-order skills of reasoning and drawing inferences, and this contributed to the growing interest in developing CT skills in K-12 and in post-secondary education. The report's most alarming findings, about a *decline* in these and other skills, have been well criticized (Guthrie 2004), to a degree that could be called "debunked". Even so, the report expressed a growing conviction that the development of students' CT skills is one of the most important functions of the education system, which might do considerably more to serve that function. A new round of lamentations occurred in 2011 when Richard Arum and Josipa Roksa published their bombshell *Academically Adrift*, which argues that most students learn almost nothing at college or university, in that they show statistically negligible improvements in the high-level intellectual skills that schools claim to teach, especially CT — though again, some of the most shocking findings have been challenged (Lederman 2013).

This concern about a general dearth of CT, at least partly attributable to failures of the K-12 and post-secondary education systems, form a backdrop to this project. CT courses are now common at colleges and universities, typically offered by philosophy departments, though taken by large numbers of students who will study no other philosophy. It's an obvious way to address the hopes and concerns just reviewed: create a course devoted to the improvement of CT skills and have lots of students take it. But this raises many questions about how to design and deliver such a course. In this project I'll explain and defend some of my answers.

1.1. Historical Background

1.1.1. Before Dewey

The term *critical thinking* and the CT movement are products of the 20th-century, beginning with John Dewey. 20th-century developments in philosophy, psychology, and education contributed to the current widespread enthusiasm for seeing that our education systems promote and enhance CT skills and dispositions. We want students to acquire the knowledge and skills of the specific fields they study in school, but also to become equipped to confront life's many problems of belief and decision. But although the concept of CT emerged in the 20th century, this aspiration has a much longer history, particularly in philosophy. Mainly that's because an important (though sometimes disputed) component of CT is informal logic, the study of arguments and fallacies, and there is a long history of philosophical attention to informal logic. The purpose of this short section is to note a few significant developments in that history in order to place my discussion of CT in a broader context.

There are examples in early Greek philosophy of what might be seen as precursors to CT. The method of demonstrating error in someone's beliefs by revealing a contradiction between one of their beliefs and an implication of their other beliefs is a reasoning skill seen in the pre-Platonic philosophers Xenophanes and Protagoras[2], and most notably Socrates. Even if the fundamental purpose is persuasion, as it is with Protagoras, such a strategy is a legitimate form of objection when it's intellectually honest — something in someone's beliefs must be wrong if these contain a contradiction. A clearer precursor to CT is Aristotle. The *Prior Analytics*, on deductive reasoning, and the *Rhetoric* and *Topics*, which examine practical and persuasive applications of reasoning, lay some of the foundations of informal logic. The *Sophistical Refutations* is a theory and catalog of the fallacies seen to be common (e.g. equivocation, composition, irrelevant conclusion), many of which remain a fixture of guides to informal logic.

A more direct line to 20th-century CT begins in the 17th century. The so-called *Port-Royal Logic* of Antoine Arnauld (written with Pierre Nicole) (1996), first published in 1662 and extensively thereafter, is a guide to the "art of thinking" that establishes some of the approaches to improving real life reasoning and communication skills found in modern CT textbooks. Part of this approach is what Maurice Finocchiaro calls "logical instrumentalism", avoiding "logic for logic's sake" (1997: 397). Arnauld and Nicole use realistic, meaningful examples from a wide variety of fields so that, they say, "the rules and the practice can be seen simultaneously" (1996: 16), and for the sake of broad appeal, engagement, and memorability. Emphasizing practical communication, they note that enthymemes (syllogisms with implicit statements) are a normal way to express arguments but also a way in which premises may escape appropriate scrutiny (1996: 176). They cover traditional fallacies from Aristotle and also several fallacies of a more psychological character — "committed in everyday life and in ordinary discourse" (1996: 203) — that anticipate contemporary perspectives on the phenomenon of motivated reasoning. They work through rules of evidence and probabilistic inference, as well as what are now called "decisions under risk" (decisions with probabilities assigned to the possible outcomes of the available actions) (1996: 273).

Several other early modern philosophers figure in the path to the 20th-century concept of CT, in particular by their treatment of fallacies. John Locke's 1690 *Essay Concerning Human*

[2] Diogenes Laertius claims that Aristotle credited Zeno of Elea with innovating the dialectical method (Laks and Most 2016: 197).

Understanding (1975) began the tradition of categorizing fallacious "*ad*" arguments (*verecundiam* [authority], *ignorantiam*, *hominem*), though he thinks of these merely as unreliable arguments, not as a specific kind of error called a *fallacy*. Isaac Watts (1726), William Duncan (1748), and Jeremy Bentham (1824) contributed further to this fallacy tradition and to the post-*Port-Royal* interest in the promotion of effective reasoning for everyday life.[3]

When Richard Whately published *Elements of Logic* (1826), he was responding to two centuries, since Francis Bacon, of skepticism about the utility of classical syllogistic logic (McKerrow 1987; Van Evra 2008). He thinks logic has been unfairly disparaged for failing to live up to misguided expectations, and he champions deduction, calling logic the "grammar of reasoning" (1826: 11), which "enables us to use the knowledge we possess to the greatest advantage" (1826: xxx). His account of fallacies construes them as either invalid arguments or arguments that are defective in some way in spite of being valid. John Stuart Mill, in his *System of Logic* (1843), does not directly challenge Whately's rehabilitation of syllogistic logic but places much greater emphasis on an empirically oriented logic of induction more closely allied to Bacon and Jeremy Bentham. For Mill, "deductive logic would impose consistency constraints on consequence-drawing, but it would not give us positive principles of reasoning" (Gabbay and Woods 2008: ix). Accordingly, Mill's account of fallacies holds them to be inferences from one or another sort of faulty or deceptive evidence.

Prior to the emergence of the concept of CT in the early 20th century, the interest in fallacies as defective argument forms is a clear precursor to CT, although fallacies have a diminished role in some contemporary understandings of what CT should be, including mine. I'll return to these issues in Parts 2 and 3. *The Port-Royal Logic* is a distant but inspiring precursor to the CT movement, and several aspects of my course have affinities to the priorities and approaches it established, such as the determination to use examples that are as realistic and meaningful as time and students' background knowledge allow, the forthright acknowledgment of the degree of implicit content in the ordinary communication of arguments, the highly instrumental treatment of logic, the inclusion of some probabilistic reasoning, and a partially psychological characterization of fallacies. There is some explicit concern with formal education in the sources above. Whately discusses the place of logic in the broader university curriculum, for instance. But it's not until the

[3] Bentham's impressive catalog of fallacies is a wonder to read just for the original fallacy names, e.g. "vaguely-insulting-vituperator's argument"; "browbeating-incapacity-confessor's, browbeating-ignoramus', or self-stultifier's, argument"; etc.

20th century that informal logic and “the art of thinking” become part of a movement anchored in education. I’ll turn to that next.

1.1.2. From Dewey to the Contemporary Critical Thinking Movement

The term *critical thinking* originates with Dewey’s earliest major work on philosophy of education, *How We Think* (1910). It’s sometimes said that he uses the term *critical thinking* in these earlier writings but later comes to favour *reflective thinking*, which appears in the subtitle of the 1933 edition of the work. In fact, from the first pages of *How We Think* (1910), he is concerned with *reflective thinking*, and it is for this concept that he gives his famous definition: “active, persistent and careful consideration of any belief or supposed form of knowledge in the light of the grounds that support it, and the further conclusions to which it tends” (2010: 6). *Critical thinking*, for Dewey, is not a synonym for *reflective thinking*, but attaches to a distinctly narrower concept. He contrasts critical thinking with “uncontrolled thinking”, by which he seems to mean epistemic promiscuity, since he claims that “the essence of critical thinking is suspended judgment” to allow “inquiry to determine the nature of the problem before proceeding to attempts at its solution” (1910: 82). This quite specific notion is not, or certainly not all of, what the term has come to mean since Dewey, although some contemporary definitions come close, such as Brookfield (2011), who sees the essence of CT as “questioning assumptions”. Since Dewey, writers have generally attached his term *critical thinking* to a concept much more like his reflective thinking.

The terms *critical* and *reflective* are used interchangeably in the 1930s by the Progressive Education Association of which Dewey was president for a time and for which his writings provided significant inspiration (Commission on the Relation of School and College of the Progressive Education Association 1943). Edward Glaser is among the first to settle on the term *critical thinking* to refer to something like Dewey’s concept of reflective thinking. He incorporates Dewey’s definition into his own definition, which distinguishes three components of CT: “(1) an attitude of being disposed to consider in a thoughtful way the problems and subjects that come within the range of one’s experiences, (2) knowledge of the methods of logical inquiry and reasoning, and (3) some skill in applying those methods” (1941: 5). Glaser’s dissertation presents an important early attempt to measure the efficacy of teaching CT, as measured in a controlled experiment using a pre-test/post-test procedure with a test of general mental ability and the newly developed Watson-Glaser CT test. He reports modest success for his own attempts at teaching CT. His most influential contribution, though, has been this definition of CT as a composite trait consisting of dispositions, knowledge, and skills.

Among the most influential contributions to modern educational theory, Bloom's Taxonomy of Educational Objectives does not employ the label *critical thinking* in the taxonomy itself, though the authors invoke *critical thinking*, *reflective thinking*, and Dewey in their discussion, which makes it clear that Glaser's "logic" and "reasoning" are less their focus than more general "intellectual abilities and skills" for problem solving (Bloom *et al.* 1956: 38). The taxonomy for cognition (a second and third taxonomy are offered for the affective and psychomotor domains) is a six-level pyramidal hierarchy with a base – Knowledge, Comprehension, and Application – supporting higher levels – Analysis, Synthesis, and Evaluation. All the levels atop Knowledge, particularly the top three, incorporate elements of what would generally be recognized as CT as well as other elements that would not be. For example, although these levels include skills such as recognizing unstated assumptions, proposing ways of testing hypotheses, and indicating logical fallacies, they also include organizing ideas in writing and "comparison of major theories about particular cultures" (Bloom *et al*. 1956: 207), activities that may involve CT, but are not plausible elements of a definition.

It would be hard to overstate the influence that Bloom's taxonomy has had within education studies and within the whole field of education. But it is not a universally celebrated contribution. Before and since it was officially revised in 2001 (Anderson *et al*. 2001), it attracted criticism of its empirical assumptions and its usefulness for curriculum design in general and for CT in particular (Paul 1985). Roland Case of The Critical Thinking Consortium argues that the nearly universal misunderstanding and misapplication of the Bloom theory of *assessment* as a theory of *teaching* has made it "one of the most destructive theories in education" (2013: 1). One of the ways it's done this, Case shows, is by encouraging teachers to assume that the higher order CT skills are more difficult than the lower order skills, and so are best introduced much later or even reserved for elite students (2013: 3). In fact, less demanding versions of tasks in the higher order "synthesis" and "evaluation" tiers, where most of the CT skills are found, can usually be devised that are suitable and beneficial for all or most students. Ironically, the use of Bloom's taxonomy may actually have stifled CT pedagogy in some classrooms even though it was one of the principal ways by which some CT skills were recognized and promoted in the decades before (and while) the CT movement gained momentum in the 1980s.

An important voice in the philosophical discussion of CT for decades, Robert Ennis did his doctoral research on a test for CT and then published an important paper proposing a pithy new definition: "the correct assessing of statements" (1962: 83). This definition, which Ennis later retired and replaced with a new one (see 1.2.1 below), is curiously lean in comparison to both earlier and later definitions, lacking in any indication of what thinking critically actually involves, what critical thinkers actually do. These details Ennis goes on to provide in a theory that distinguishes logical, criterial, and pragmatic thinking. Perhaps the most important feature of Ennis' 1962 definition is that it makes CT a normative concept: there are methods of assessing statements that are correct, and to think critically is to use those methods. Ennis claims that he's simply acknowledging "the predominant manner of speaking" by "build[ing] the notion of correct thinking into the notion of critical thinking" (1964: 599), although this manner of speaking is not reflected in the earlier definitions he cites (Smith 1953) or in Dewey's definition. Glaser's definition hints at this. His "methods of logical inquiry and reasoning" and "skill in applying [them]" gesture towards standards of good thinking. This normative aspect seems to establish a continuity with the pre-Dewey forerunners to CT, for example in the attention to the normative concept of a fallacy. During the period of the progressive education movement, with its enthusiasm for student-centred educational reforms, the emphasis was on reflective, exploratory, and self-regulatory thinking behaviour more than on adherence to principles of correct thinking.

To summarize: Throughout this period, we see the concept of CT emerge as a key component of progressive education; the terminology stabilizes and the concept becomes a freestanding educational goal with a dispositions-knowledge-skills structure; hallmark CT skills are incorporated into highly influential educational theory; and an important proposal affirms anew the normative character of CT.

1.2. Defining CT

1.2.1. Contemporary Definitions

Definitions of CT have proliferated since the 1980s when the contemporary CT movement began and it became common, even *de rigueur*, for institutions of higher learning to declare their commitment to promoting CT and building the dispositions and skills of CT. I will make no attempt to catalog and assess all the published proposals. Within these proposals, even the existence of major disagreement is a point of disagreement. Some authors see a broad consensus rooted in Glaser's definition (Haber 2020: 37). Others survey the numerous proposals on offer and see "widespread disagreement" about the nature and purpose of CT (Tiruneh *et al*. 2016: 484), with the concept actually becoming "less, rather than more, clearly defined" with decades of scholarly attention (Capossela 1998).

There is considerable diversity, which I sample below, of philosophical views about what it means to think critically, but there are indeed points of consensus. I will lay out my own understanding of the critical thinker in the next section. This section will fill out some of the more recent history and context of my project. This will help to explain my decision to sidestep the problem of formulating a traditional short definition and to proceed directly to a picture of the critical thinker with enough detail that it doesn't immediately demand further definitions or explications of its key terms. I will not advance yet another competing definition of CT – this is not a point on which more innovation is needed. I'll articulate a vision that is generally within the mainstream, offer some justifications for it, and discuss its implementation in the classroom. The last thing I'll do in this section is identify a few aspects of the cited definitions that I think are particularly important to incorporate into my picture of the critical thinker.

The surge of interest in CT in the 1980s saw numerous proposals about the nature of CT from psychology, philosophy, and education. The psychologist Robert Sternberg characterizes CT as "the mental processes, strategies and representations people use to solve problems, make decisions and learn new concepts" (1986: 3). Notably absent from a psychological definition like this is any reference to logic and rationality or even to standards and criteria. Sternberg sees these concerns as characteristic of a philosophical approach to CT that is of limited real-world relevance, providing an account only of how people think "under ideal circumstances in which the limitations typically placed upon the human information processing system are not in place" (1986: 5). It's unclear whether Stenberg's problem solving and decision making include solving problems of, or making decisions about, what or how strongly to believe.

That question is answered directly, and affirmatively, by Ennis in his updated definition: CT is "reasonable reflective thinking that is focused on deciding what to believe and do" (1987: 10). The new definition is again a normative one, if slightly less bluntly so, as are Harvey Siegel's claims that the critical thinker is someone who is "appropriately moved by reasons", and that CT is the ability to "assess reasons properly" along with the "disposition to base one's actions and beliefs" on the assessed reasons (1988: 25). So also is Matthew Lipman's proposal, which he claims corrects Sternberg's too broad definition and Ennis' too narrow definition: CT is "defined by three characteristics: 1) it is *self-corrective*; 2) it is thinking with *criteria*; and 3) it is thinking that is *sensitive to context*" (1987: 5). All three characteristics immediately call for explication. These and similar short definitions can only anchor some larger vision or theory of CT.

In the same productive period as the definitions above, Michael Scriven and Richard Paul offer a more comprehensive (several paragraphs) definition that stands on its own. The core of it is this:

> Critical thinking is the intellectually disciplined process of actively and skillfully conceptualizing, applying, analyzing, synthesizing, and/or evaluating information gathered from, or generated by, observation, experience, reflection, reasoning, or communication, as a guide to belief and action. In its exemplary form, it is based on universal intellectual values that transcend subject matter divisions: clarity, accuracy, precision, consistency, relevance, sound evidence, good reasons, depth, breadth, and fairness… Critical thinking can be seen as having two components: 1) a set of information and belief generating and processing skills, and 2) the habit, based on intellectual commitment, of using those skills to guide behavior. (Scriven and Paul 1987)

This active period of theorizing about CT culminated in the best known long format definition, the Delphi consensus definition, produced by a large panel for the American Philosophical Association in 1990:

> We understand critical thinking to be purposeful, self-regulatory judgment which results in interpretation, analysis, evaluation, and inference, as well as explanation of the evidential, conceptual, methodological, criteriological, or contextual considerations upon which that judgment is based… The ideal critical thinker is habitually inquisitive, well-informed, trustful of reason, open-minded, flexible, fair-minded in evaluation, honest in facing personal biases, prudent in making judgments, willing to reconsider, clear about issues, orderly in complex matters, diligent in seeking relevant information, reasonable in the selection of criteria, focused in inquiry, and persistent in seeking results which are as precise as the subject and the circumstances of inquiry permit. (Facione 1990: 2)

What's immediately clear about the short, dictionary style definitions of CT is that they are barely useful as an answer to the question "What is critical thinking?". The concept of CT is so multifaceted that anyone who wants to understand what CT is, never mind to arrange a curriculum for the enhancement of CT, will immediately have to refer to the subsequent elaboration of the definition. The impressively dense Delphi definition, and the longer Scriven and Paul definition, succeed at being more or less free-standing. They are useful enough on their own that a quarter century after its formulation, the Delphi definition was deemed suitable for guiding an ambitious meta-analysis investigating the effectiveness of CT teaching strategies (Abrami 2015). These long format definitions underscore the fact that someone who aims to improve students' CT could do so by improving many different aspects of their thinking.

My task in the next section is to present a picture of the critical thinker. It will emphasize some elements of these definitions and de-emphasize others. Rather than formulate a short definition or slogan that cries out for elaboration, or a dense definition that attempts to pack an adequate account of an expansive concept into a couple of relentless sentences, I'll offer a reasonably self-contained, and so necessarily much longer, account of CT.

From the above definitions I'll extract not particular CT traits but rather some general features of a good definition itself.

1) CT is best defined by describing both a type of activity and a type of person.

The short definitions simply define CT. Scriven and Paul push through several paragraphs defining CT itself before moving to some consideration of the critical thinker, who inevitably manifests CT in incomplete ways and who therefore makes the improvement of their CT a lifelong project. Though his definition is of CT, when it comes to elaborating that definition, Ennis immediately takes his task to be "a characterization of the ideal critical thinker" (1991: 8), as does the Delphi definition. Defining CT in part by defining the critical thinker is not strictly necessary but it is a natural way to acknowledge the widespread hope that our educational systems will produce a certain kind of citizen.

2) CT has both skill and disposition components.

This aspect of CT was established early on by Glaser, whose definition informs the thinking of many contemporary writers (Haber 2020; Turnbull 2017). And we see it in the influential definitions cited above. For example, the account Ennis (1991) builds around his definition includes 16 abilities and 12 dispositions. The first of these dispositions is clarity, which also appears in Scriven and Paul's list of "universal intellectual values" upon which CT "in its exemplary form" is based. They go on to explicitly recognize "habits, based on intellectual commitments" to using appropriate thinking skills. The Delphi definition also recognizes "habits" of the ideal critical thinker, of which, again, clarity is the first. *Dispositions*, *values*, *habits*, and *commitments* are not synonyms, but these are all ways of describing the critical thinker in terms of motivations and tendencies, not just pure abilities.[4] It's not enough to be *able* to think in certain ways — the critical thinker is inclined or strives to think in these ways.

---

[4] An interesting question that I will not pursue beyond this note is whether there is really any such thing as a "pure ability" with no dispositional component. Perhaps we need to invoke dispositions to understand what an ability or skill is. Indeed, perhaps ultimately an ability can be entirely analyzed in terms of dispositions. Whether this makes sense might depend on how dispositions themselves are to be understood. What it ultimately means to say that I have the ability, for example, to identify the conclusion

A critical thinker is someone who has certain dispositions. But is it just that for every CT skill, there is a corresponding disposition "to use that skill", or a single master disposition "to use one's CT skills"? If so, it may be an uninteresting sense in which we need to incorporate dispositions into the definition. None of the authors cited seem to view the matter that simply, though. For example, Ennis' 12th CT disposition is "to use one's critical thinking abilities", whereas the other 11 — things like #4 "to seek and offer reasons" — are not helpfully characterized as simply the inclination to use a skill. I think the dispositional components of CT must be recognized in both forms, as motivations to use CT skills, and as holding or pursuing certain CT values. I'll take this approach in the next section. However just because CT has dispositional components does not automatically mean that a pedagogical intervention, e.g. a standalone CT course, should or can try to *teach* those dispositions. I'll return to this issue in section 2.1.2.

3) <u>CT is for both forming beliefs and making decisions.</u>

The Ennis definition is the most prominent example of a definition that clearly and directly emphasizes this dual purpose of CT. Scriven and Paul see CT as guiding "belief and action". The Sternberg definition has CT to "make decisions and learn new concepts", though as I say above, it's not clear whether this includes deciding what to believe. In contrast, the Delphi definition has a clear epistemic rather than practical character, with no direct reference to decision or action. Much of the real and perceived social value of CT is based on expectations about how critical thinkers will act and the quality of their decisions. I see CT affecting how people act mostly by affecting how they form beliefs and how they make use of information they believe, but the explicit inclusion of decision making in the definition has the virtue of directly acknowledging this common basis for valuing CT.

4) <u>CT is a normative concept.</u>

An adequate definition of CT is not simply a psychologist's description of a certain kind of cognitive activity. To call someone a critical thinker is to say that their thinking is reasonable or appropriate or successful by some criteria. Most of the definitions above incorporate some normative element

---

of an argument written in paragraph form might be that I have some set of dispositions. These might be things like "often when confronted with a passage that seems persuasive in tone, I search for inference indicators such as 'therefore'" to determine what claim it asserts and supports with reasons, and so on. Some of these dispositions might even be sub-personal, e.g. described in terms of motor functions. If the right and enough of these dispositions come together, this view would say, I just do have the ability to identify the conclusion of an argument. Thanks to Bruno Guindon for raising this issue.

that corresponds to the general public's sense that CT is good by definition, whatever exactly that definition is. Mine does, as well.

1.2.2. The Critical Thinker

A critical thinker:

- Values truth for its own sake and as a basis for making effective decisions.
- Seeks evidence in order to judge truths and estimate probabilities, and strives to form beliefs or credences that are justified by evidence.
- Monitors their own beliefs or credences and strives to make these coherent (consistent with one another).
- Confronts an argument:
    - As both an opportunity to gain knowledge and a hazard of being misled.
    - To identify a conclusion, thesis, or hypothesis, and to distinguish this from premises, assumptions, or evidence claims, explicit and implicit.
    - To consider as distinct evaluative questions:
        - Whether premises or evidence claims are true or probable.
        - How well these claims support the conclusion.
    - With a sense of its evidential burden, given the plausibility of its conclusion.
    - With an understanding of the style or type of reasoning of the argument, and of any errors, oversights, biases, or fallacies to which that type or reasoning is especially vulnerable.
- Follows and constructs chains of inference.
- Acknowledges the natural human tendency of motivated reasoning, recognizing it in others and in themselves in order to account or compensate for it to the extent possible.
- Appreciates the essential role of expert opinion and consensus in a world of knowledge specialization, and identifies genuine expertise.
- Treats ethical judgements as claims that demand the support of reason and evidence.
- Thinks logically, when appropriate, by:
    - Using the natural language expressions of logical concepts correctly.

    - Following rules of deductive inference, in particular in the construction, use, and critique of categorical and conditional reasoning.

- Thinks probabilistically, when appropriate, by:

    - Treating events and statements as more or less likely, and assigning probabilities in reasonable ways.

    - Making judgements and decisions that reflect estimated or calculated probabilities, that correctly apply rules of probability, and that consider the values of possible outcomes.

- Thinks scientifically, when appropriate, by:

    - Reasoning by explanation and judging the merits of competing explanations.

    - Searching for alternative explanations and remaining mindful of the difficulty of thinking of relevant possibilities.

    - Thinking of checkable predictions that are capable of confirming or disconfirming a hypothesis.

    - Reasoning inductively, understanding when generalizations or projections from limited data are either warranted or likely to mislead.

    - Articulating the structure of evidence that could demonstrate the presence or absence of a causal relationship, and perceiving when such evidence is inadequate or incomplete.

- In communication generally, and especially in a conflict of opinions:

    - Expresses their thoughts clearly and precisely, and seeks clarity and precision from others.

    - Strives to establish mutual understanding of points of agreement and disagreement.

    - Promotes productive discussion by representing others' opinions and arguments with reasonable charity.

With this picture of the critical thinker, I don't allow CT to become synonymous with good thinking generally, to encompass every desirable cognitive trait. In particular, this account makes no explicit mention of creativity. A division between critical and creative thinking is sometimes seen as fundamental. For example, the American not-for-profit educational organization Battelle for Kids promotes "4 C's" of 21st-century learning: CT, Communication, Collaboration, and Creativity. The same fundamental distinction is recognized in British Columbia's K-12 curriculum "core competencies": Communication, Thinking, and Personal and Social. Thinking is divided into just two categories: Creative, and Critical and Reflective (BC's Curriculum 2022). Within STEM education, creativity and CT are increasingly recognized as an important pair of interrelated yet distinct competencies (Ellerton and Kelly 2022). This creative-critical distinction is more plausible for some kinds of creativity than for others. It's most plausible for artistic creativity and somewhat less so for creativity in problem solving and innovation. One area where overlap seems clear is creativity in scientific thinking: formulating alternative hypotheses, predictions, and tests. Some authors already cited understand an element of creativity to be implicit in their definition of CT at exactly this place (Ennis (2013: 7), following Bailin (1999b: 288)). I see an element of creativity implicit in the corresponding parts of my account even though I exclude creativity more generally.

A definition of CT could also place greater emphasis on a cultivated instinct to question authority, power structures, mass and social media, tradition, or the status quo. I have a few elements related to these, such as demanding evidence and appreciating the role of expertise and expert consensus. But I don't offer anything as radical in this way as, for example, James Marshall (2001), who calls for a Foucauldean CT of social and political liberation, or anything that would fall under Ennis' category of "politically-motivated definitions" (2013: 8). While CT can contribute to empowering citizens and strengthening democracy, again I resist the temptation to make CT exhaust the high-level intellectual traits that we may look to our educational systems to foster in maturing thinkers. There are many desirable intellectual traits that are adjacent or related to CT but not components of it. Efforts to foster these traits may reinforce or benefit from efforts to foster critical thinking.

What I have in my account of the critical thinker is not simply a list of course learning objectives, but a picture that informs my objectives. It is pedagogically ambitious. We must be realistic about what can be achieved in a single-term, large-enrollment introductory course. One might suppose that such a course ought to give students the best all-around introduction to CT possible, where "all around" means something like "emphasizing the major aspects of critical thinking roughly equally". But this may not be an optimal strategy. There may be elements of this account that are less well suited to instruction in a standalone course, or to any instruction. The definition of CT is one thing, the optimal design of a standalone course another. A CT course is well designed if it helps students make meaningful gains in many of these traits.

1.2.3. The Combativeness Objection

That CT is a good thing normally goes without saying among scholars, educators, and the public. Many definitions of CT recognize CT to be thinking that fulfills some standards of adequacy, and the nearly universal view is that more people should do more CT. Explicit opposition to CT is seen as outlandish; even a local (apparent) instance can attract national ridicule.[5] However there is a serious criticism of CT within academia, which I consider in this section on defining CT because the objection claims that CT is inherently problematic. The concern, which I'll call the combativeness objection, claims that CT encourages, even valorizes, the undesirable traits of intellectual combativeness, aggression, and arguing for the sake of "winning". More or less this same allegation has been made about academic philosophy generally. Since CT is typically taught by philosophers, the objection amounts to a concern that contemporary CT has taken on the negative character of the discipline from which it has most directly emerged. One (the fifth) of the questions I set for this project is how the present CT course functions as a philosophy course and how it represents philosophy to students with other academic pursuits and to potential philosophy students. Insofar as the activities of this course are philosophical activities, I want them to demonstrate and involve students in philosophy at its best. While I do not grant that academic philosophy generally or CT specifically has a combativeness problem, I also don't think the concern is completely unfounded. Here I'll look at the charge, consider whether it's a serious problem with CT, and highlight some aspects of the present course that allow students to develop thinking that is appropriately critical but not combative.

A version of the combativeness objection is raised as a general problem in academic philosophy by Janice Moulton (1983), who criticizes the "adversary method" of philosophy on several grounds. She argues that the marginalization of thinkers not easily cast as intellectual combatants produces an unduly narrow view of what philosophy is, has been, and can be, that aggressiveness in the conduct and evaluation of philosophy is disproportionately alienating to women, and that the adversarial method is just not particularly good for the discovery of truth, in part because it leads to distorted and less than charitable interpretations of others' positions.

Sharon Bailin clarifies a set of similar concerns specifically about CT, including that it is "aggressive and confrontational rather than collegial and collaborative" (1995: 191), and perhaps

---

[5] The 2012 platform of the Texas Republican Party provoked widespread derision after it was noticed (Jaschik 2012) that it appeared to explicitly oppose CT: "We oppose the teaching of Higher Order Thinking Skills (HOTS) (values clarification), critical thinking skills and similar programs that are simply a relabeling of Outcome-Based Education (OBE) (mastery learning), which focus on behavior modification and have the purpose of challenging the student's fixed beliefs and undermining parental authority." As the passage indicates, though, it was not really CT itself that the party claimed to oppose, but an educational model called Outcome-Based Education, which the party viewed as a vehicle for advancing certain liberal causes. The Republicans were still much in favour of what they called CT for "controversial" topics such as biological evolution and anthropogenic climate change (Owen 2012).

thereby biased against whole groups of thinkers who operate in these less valued ways, although she does not herself advance the claim that CT is biased in this way. Barbara Thayer-Bacon (1998) takes a more critical stance. She wants to "transform" CT into "constructive thinking" by more firmly establishing a place for the "tools of imagination, intuition, and emotional feelings" and by recognizing and embracing the interactive (as opposed to individualistic) nature of knowledge production. In a related charge, Michael Roth (2010; 2021a; 2021b) sees in CT a closed-mindedness and combativeness that inhibit the sort of exploratory, empathetic conversation and collaboration that society needs more of. Denouncing the way CT is "fetishized" by people in education (2021a), Roth's rationale for urging more "critical feeling" and less critical thinking is a concern about the political manipulation of emotions, in particular the gratifying but destructive emotions of righteousness and resentment. "Critical thinking alone will not turn us from such pleasures; reason alone never supplants sentiment. We need critical feeling — practiced emotional alternatives to the satisfactions of outrage" (2021b). And David Hayes doesn't mince words when he declares that CT "is seriously misguided as an educational goal and the time has come to find another pedagogy" (2015: 318). The core of the problem is that CT pedagogy "tends toward aggression" and its "mode of relation to an object… seeks and uses the emotions appropriate to conflict" (2015: 319-320). When in this attack mode, he claims, critical thinkers make a "premature leap" to assumptions about the content of a critical target (person, text, or idea). "The critical thinking mode tends to see only one ethical-intellectual virtue: courage", often taking the form of an aggression that is "especially undermining" of classroom discussion (2015: 321). When students focus on tearing down beliefs and justifications, they end up taking no position on anything.

The claim that CT has some element of intellectual contest or combativeness is not completely unfounded. As in philosophy generally (Rooney 2010: 211), we can glimpse this in the language of CT. One could point to Ralph Henry Johnson and Anthony Blair's landmark 1977 textbook (2006), titled *Logical Self-Defense*; it has a long tour of fallacies categorized as attempts at diversion, intimidation, etc. Ann Cahill and Stephen Bloch-Schulman (2012) develop a method of learning CT that is explicitly modeled on martial arts pedagogy. Outside of the academy, Kevin deLaplante (www.argumentninja.com) promises to turn people into "argument ninjas" with CT and persuasion skills. These are not just superficial promotional choices, either. It really is one goal of mainstream CT, including my course, to equip students to defend themselves from error and deception, and such defence does involve noticing and criticizing flaws in attempted persuasion.

If the combativeness objection is right that CT encourages incivility and criticism for its own sake, or pointlessly turns interactions that could be collaborative projects of truth discovery into intellectual jousting, or banishes sympathy and empathy from attempts to resolve disagreement, CT is indeed flawed and the strong push to promote it is a mistake. It's senseless to deny that the undesirable features named in the combativeness objection are ever exhibited by people who take themselves to be thinking critically. I claim, first, that the objection does not identify any problem intrinsic to CT, and second, that there are several aspects of the present course that mitigate the risk of encouraging combative traits in students.

In Section 2.1.2, I'll say more about the limited role and ability of a single CT course in teaching or instilling values, but Hayes' claim that courage is the sole virtue prized by CT again flies in the face of what mainstream CT strives for. For example, in the list of intellectual virtues of the California-based Foundation for Critical Thinking, courage does indeed appear – along with humility, empathy, autonomy, integrity (honest arguing), perseverance (to hard thinking), confidence in reason, and fairmindedness. It doesn't appear in the APA or my own definitions at all. If Hayes claims that toxic courage is the true character of CT, he owes far more evidence that a project with these explicitly endorsed virtues somehow ends up honouring only courage-aggressiveness. Moreover his complaints that CT encourages and even rewards intellectual vices like straw-manning and uncharitable interpretation (2015: 321) are exasperating. These very concepts are, of course, classic elements of the vocabulary of CT. A comprehensive CT education tries to help people notice and avoid these very unfortunate tendencies. Hayes' own criticisms appear to exhibit these vices. In any case, his very use of these concepts is an implicit endorsement of CT.

As Trudy Govier (1999) argues, a certain "minimal adversariality" (adversariality that adheres to norms of civility and politeness) does play an essential role in CT, particularly in political contexts. (I'll return to this idea in Part 3.) Undoubtedly politeness norms can fail, and perhaps there are other (e.g. feminist) reasons to be concerned about these norms even if they are reliable (Hundleby 2019), but minimal adversariality is not something that CT should apologize for. Minimal adversariality means understanding that in arguing for ***P***, one is committing to arguing against NOT-***P***, which will be at least a hypothetical and perhaps an actual position held by others. But this stance need not "slip" into belligerence or hostility toward a person who holds that NOT-***P*** (Govier 2021: 526-528). We should want our students to comprehend that holding a view entails taking a position on the possible beliefs of others, and that disagreeing with those others can not only be done respectfully but can be a *way* of respecting them, by treating their view as worthy of engagement and well articulated objection. It eschews a lazy relativism that doesn't take the other's view fully seriously. Along with modelling of what respectful disagreement looks like, this form of adversariality can encourage epistemic humility, since it underscores that justification for a belief involves some burden to be able to articulate a criticism of its negation. Of course epistemic humility need not and should not become a general reluctance to take any positive position on anything, as Hayes worries. But as Hayes makes clear, this is an outcome to be concerned about with unchecked "ancillary adversariality" (Govier's term), CT as hunting and attacking the beliefs of others.

A slip from minimal to ancillary adversariality, from a critical to a combative stance, is obviously possible, and we should be aware that CT education might provoke it in certain personality types predisposed to a warrior mentality – which may well, as some critics fear, appear disproportionately in male students. And if there are CT instructors, textbooks, or courses that encourage this, these do a disservice to students and to society. But this is not an intrinsic feature of CT. But the flaw here is in the interpretation or implementation rather than in CT itself. This does not represent CT as I, or as the APA or Ennis, *et al*., understand it or strive to practice or promote it. (I won't speak for the Argument Ninja.) Now this response could invite a charge of the "no true Scotsman" fallacy: faced with possible counterexamples, I'm falling back on "true CT". One reason I don't teach this fallacy is that I think it's rarely a fair characterization of a conceptual disagreement. There is a mainstream set of overlapping, thoughtfully developed definitions, including mine. These articulate in detail what CT is and what good CT looks like, and according to them, if the teaching or practice of CT sees productive adversariality slide into unproductive combativeness, this falls short of thinking critically. Perhaps someone observing this thinks that it does reveal the true character of CT, and so here's a problem with CT. Then we have a disagreement about the concept of CT. My understanding of CT could be wrong, but the disagreement isn't resolved by crying "Fallacy!". My view is that such unfortunate cases would be a reason to do CT better, not to reject CT as a flawed ideal.

Since CT education *could* encourage combativeness if done poorly, it's important that the present course include some reasonable precautions not to, both for the sake of its quality as a CT course and to not contribute to a larger "branding problem" with philosophy – with those who might be invited into philosophy through the course and with non-philosophers about whose impressions of the discipline we should care.

One way in which a CT course could inadvertently encourage combativeness in students is by training them to view their information environment as a minefield of errors and fallacies to be detected and destroyed. From the outset, I present arguments and scientific reasoning as opportunities to learn and ways to organize and communicate reasons. Certainly these bring with them hazards of being led into error: bad arguments, motivated reasoning, poorly considered or utilized probabilities. But I teach these things as part of teaching what good reasoning is. After the introductory Unit 1, three units present arguments as attempts to convince but not as attempts to deceive. Arguments are reasoning structures that can be better or worse by objective standards, but students don't learn to approach every argument as a potential fallacy or deception. Learning about invalidity here is primarily a way for students to better understand the special character and requirements of valid inferences. (Similarly, in the section of Unit 2 on counterexamples, the point is not to develop a technique for shooting down generalizations but to better understand what a generalization denies and what it allows.) Students don't learn invalidity as a fatal flaw to spot in others' arguments. I think that clear real-life examples of invalid deductive arguments are not as common as some treatments of deductive arguments suggest, so training students to spot and dismiss invalid arguments, a standard CT exercise, does create unnecessary risk of encouraging combativeness.

Some arguments that might be diagnosed as invalid plausibly contain an implicit premise that would make them valid, though whether such an implicit statement is "really" present might be indeterminate, as the author may be less than fully clear on their own commitments unless directly prompted to make them fully explicit. Training students to denounce such an argument as invalid rather than to take a more nuanced view of people's language and beliefs would risk encouraging combativeness, so the validity exercises in Unit 2, for example, are exercises in articulating the non-truth-preserving nature of an invalid inference and constructing inferences that are truth-preserving.

And some arguments might be inappropriately diagnosed as invalid by being inappropriately treated as deductive. In particular, an affirming the consequent "fallacy" is usually more charitably interpreted as a (non-deductive) confirmation argument. In the Unit 2 class where we work through an example of an apparently invalid argument (the example discussed in note 21), the lesson is that invalidity is real and understanding it is a necessary component of understanding validity, but there is a difference between *invalid* and *unreasonable*. An argument that is formally invalid is *not* thereby a bad argument, but only bad by one special and demanding standard.

Units 6-7 are devoted to reasoning with the tools of mathematical probability. They point out a few seductive errors (e.g. the gambler's fallacy, the conjunction fallacy, base rate neglect) but don't drill students on attacking them. By using subjective probability as a way of describing degrees of belief, both others' (in examples and questions) and their own (in Assignment 2), students learn to conceptualize epistemic options between the dogmatism or extreme confidence that could be seen as confrontational and the general skepticism or disinclination to take any position that Hayes sees as a common student reaction in an environment of intellectual combativeness.

To be sure, the course certainly does incorporate some criticism exercises, in particular in Units 5 and 8-10, where students practice identifying and describing certain kinds of biases and weak arguments. But these are tied directly to material whose primary function is to help them recognize, comprehend, and construct examples of specific types of inductive reasoning (confirmation, analogical, generalizing, and causal). For example, to truly understand what a good sample is, students should be able to articulate what exactly is wrong with a biased sample, something that turns out to be more challenging that one might expect. Since this is not a statistics course, a qualitative explanation of a biased sample, or a sample of size insufficient to support a particular inference, is a useful way to allow students to demonstrate their understanding without learning CT as a habit of scrapping or sniping.

1.3. The Generality of Critical Thinking

1.3.1. The Domain-Generality Question

The first in the set of questions I posed for this project in the introduction asks why, even given that society is looking to formal education as key to the early adulthood development of CT, CT should be the subject of a standalone course, a course dedicated to CT in general. The most pressing aspect of this question is whether such a standalone course might be fundamentally ill-conceived, notwithstanding the widespread offerings that go by that title. This project concerns the design of a one-term (trimester) CT course taught at the first-year level of college and university. To create and offer this course, as well as the many other similarly positioned courses, presupposes that CT can usefully be taught in this format. That is, a course that treats CT as its own academic subject is based on the assumption that CT skills are to a significant degree subject- or domain-general, such that they can be taught in a generalized way, for example as exercises of informal logic. The objective is to enable students to recognize the applicability of these skills and to become able and inclined to apply them in other academic and life contexts. But if CT skills are highly domain-specific, such that there really are no principles or skills of "CT in general", but only "CT about subject X", "CT about subject Y", etc., then a CT course is futile in principle. CT could be taught only as an aspect of, not as principles applied to, the content of various subjects. Hence, whether CT is domain-general or domain-specific has significant implications for the optimal design of CT education. And if the optimal design is not feasible, this may still have implications for whatever sub-optimal design is feasible. In this section, I review and draw some conclusions about this important issue, and I consider some implications for the design of a standalone CT course.

Ennis distinguishes two forms of the domain-specificity thesis: 1) conceptual domain-specificity, the claim that "there are no non-trivial general principles of critical thinking at all"; and 2), empirical domain-specificity, the claim that "critical thinking principles learned in one situation will as a matter of empirical fact not transfer from that situation to another type of situation" (2013: 11). While this nicely summarizes two distinct contentions, it doesn't really serve to categorize specifist authors since most make both claims to some degree, the conceptual claim providing a partial explanation of the empirical claim. It also doesn't directly convey anything about the alternative terms of the debate, skill transfer. The generality question is closely linked to but distinct from the question of transferability, whether a principle or skill learned in one context can be successfully abstracted from that context and applied in an appropriate way in other contexts. If it is anchored to or embedded in a specific domain – it cannot be transferred – then it is domain-specific. Generality and transferability are not the same concepts, though. The generality of a skill enables it to be transferred. By definition, a domain-specific skill can be taught and learned only within and with reference to a particular domain, whereas a general one could in principle be taught and learned either in a particular domain, and then transferred between domains, or in general terms, and then applied to particular domains. I'll return to this point below.

Perhaps surprisingly, given that there is no indication of it in his famous definition of CT, the specifist view goes back at least to Glaser: "Development of skill in applying the methods of logical inquiry and reasoning… appears to be specifically related to, and in fact limited by, the acquisition of pertinent knowledge and facts concerning the problem or subject matter toward which the thinking is to be directed" (1941: 175). In this and his later published views (1984), he advances a version of empirical specifism, as he doesn't deny or diminish "the methods of logical inquiry and reasoning". Bailin *et al.* take a similar view, seeing CT skills as bound up with knowledge of particular subjects: "Skilled performance at thinking tasks cannot be separated from knowledge.

The kinds of acts, such as predicting and interpreting, which are put forth as generic skills will, in fact, vary greatly depending on the context, and this difference is connected with the different kinds of knowledge and understanding necessary for successful completion of the particular task" (1999a: 271).

Daniel Willingham also endorses empirical specifism: "There are indeed principles that carry across domains of study. The problem is that people who learn these broadly applicable principles in one situation often fail to apply them in a new situation" — though a few paragraphs later he edges toward the conceptual view: "We're not even sure the general skills exist, but we're quite sure there's no proven way to teach them directly. In contrast, we have a pretty good idea of how to teach students the more specific critical thinking skills" (2020: 42-44). He reviews some dispiriting evidence of people's inability to notice and exploit the underlying structures of problems, for example in being no better at solving a problem even when they have just seen and understood the solution to a directly analogous problem. I can attest to this phenomenon, as I'm sure many instructors in many fields can. (See Appendix D.) The contention here is that CT can be developed only within specific domains and depends on deep knowledge of those domains. The core obstacle for generalized CT education is the limited transfer of skills learned in one context to different contexts, and instruction in generic CT skills lies somewhere between "ineffective" and "meaningless" (Hendrick 2016).[6]

The arch-specifist is John McPeck (1981), who comes closest to holding conceptual specifism: "There are almost as many different kinds of critical thinking as there are different kinds of things to think about. The criteria for applying and assessing critical thinking derive from the thing (call it a topic, subject, field, or domain) being discussed or thought about" (1990: 10). The core of his argument is: "Because objects of thought can and do differ enormously in scope, quality, and variety, I argue that there can be no one general skill or limited set of skills (including formal logic) which could do justice to this wide variety of objects" (1990: 11). He's also pessimistic about skill transfer (1990: 12). However even McPeck does not deny the existence of all general thinking skills. He just thinks that their usefulness is inversely proportional to their generality. Acknowledging highly general — but useless! — CT principles appears to be how he makes it

---

[6] Wellingham confirms in a note that the problem for generalism is not its conceptual emptiness or uselessness but rather people's cognitive limitations. He concedes that CT really does exist as a set of general skills that can be flexibly applied in new contexts… by smart people, "people with extensive training, such as Ph.D-level scientists" (2010: 26). Peter Ellerton (2022) discusses and critiques Wellingham's empirical pessimism, which he treats as an example of *cognitive load theory*, whose key pedagogical recommendation is the careful management of the highly limited resource of working memory.

possible for specifists to even speak of one thing, *critical thinking*, that students might develop in different domains or about different objects of thought. Of course, as critics of specifism readily point out (Smith 2002), this "one thing" can and should still be conceived of as a set of skills, not literally one skill, a single method or procedure called *critical thinking*.

The domain-specific view of CT doesn't imply that CT cannot be taught, but as I say above, it does imply that a standalone CT course would be pointless. This view is widely contested, not only by those, like Ennis (2013) and Siegel (1988), who directly defend generalism, but also implicitly by the many philosophers who don't believe they're wasting their and their students' time by teaching and creating such courses. I count myself among these. Specifists' arguments fall well short of establishing that there are no general principles of CT worth teaching or learning. For example, McPeck's observation that objects of thought differ in many ways can hardly be disputed. So indeed, in some ways, CT has to be a somewhat different exercise when it's about different things: the authorship of the plays of Shakespeare, the QAnon conspiracy theory, the Russian invasion of Ukraine, the Atkins diet, etc. Even for any one object of thought, different questions, goals, and types of evidence mean that the exercise of CT could take very different forms. But the fact that every application of CT is unique in some ways does not mean that there is no "limited set of skills" that are a legitimate and useful object of study.

At the beginning of the course, I introduce CT skills with the analogy of a set of tools. A handyman goes to a job with a toolbox containing a limited set of tools. Part of what it is to be a good handyman is to have and be proficient with a set of tools, and to understand what a successful application of each tool would be with a large range of problems. If my toolbox contains only a hammer, or if it contains a hammer and a screwdriver but I try to use the hammer on a screw, or if I turn my screwdriver counterclockwise to drive the screw down, or if I think that the screw head left protruding well out from the wood surface is well inserted, then I am a bad handyman. Indeed, if I go around to different jobs and discover that each problem is unique, as will inevitably be the case, and conclude that it was a mistake to purchase my toolbox with its limited set of tools, I have not gained a deep insight into the nature of handymanning. On the contrary, I am simply not yet a competent handyman.

Becoming a competent critical thinker is partly a matter of gaining a reasonable number of skills that are each applicable to a reasonable range of problems (of comprehension, communication, belief, decision), and of being able to recognize which skill is appropriate and apply it effectively. For example, a critical thinker should be able to conceive of an explanation as a hypothesis that may be tested by generating and checking a prediction against a background of assumptions (auxiliary hypotheses). This is not a master thinking skill that is the key to every successful episode that calls for CT, but it is a general skill that is applicable to a wide range of scenarios in which one might want to decide what to believe or how to adjust one's credences. Contrary to the conceptual specifist, the generality of this skill does not render it useless. To recognize, for example, that a hypothesis test may take the form of a prediction, and that any prediction for any hypothesis is based not only on that hypothesis but also on other assumptions, gives one an important advantage in a *new* situation in which one must determine an appropriate degree of belief. Most of the specific background assumptions that are worth recognizing (being less than certain) will be completely different for different hypotheses and in different domains. But understanding this general feature of predictions without having to separately notice of each particular prediction that it is also based on its own particular assumptions, so that one immediately knows to start thinking of assumptions worth questioning in any new predictive situation, is a clear advantage, not some empty CT slogan.

Ennis' long-running complaint (1989) about subject- or domain-specificity is that *subject* and *domain* are too vague for the thesis to be meaningful. McPeck (1990) tries to deflect this criticism with the rebuttal that if the terms are vague, that is equally a problem for subject- or domain-generality. But this response falls flat. The generalist does not need to specify some domain that CT skills are "general to". In this sense, generalism is a negative claim: CT skills are *not specific* to particular domains, subjects, or objects of thought. It's also never made clear why specifism would afflict only CT. For example, a standard course in English departments may be called something like "Principles of Literary Criticism". In some sense, literary criticism "cannot be separated from knowledge [of works of literature]. The kinds of acts… which are put forth as generic skills will, in fact, vary greatly depending on the context". But are these lit-crit courses thereby pointless? The specifists don't seem to be saying so. And are there "as many different kinds of" evolutionary biological thinking or Keynesian economic thinking as there are different biological or economic objects of thought? Specifism seems to contain no principle that would limit it to CT, yet it's clearly implausible of other subjects.

The empirical case for domain-specificity seems to conflate two issues. In fact it appears to be a bit of a bait-and-switch. There are indeed some reasons to be pessimistic about whether most people can acquire CT skills in one domain and spontaneously transfer those skills to other domains, whatever exactly *domain* means. However this interdomain transfer is not what generalism claims is possible or recommends (though it doesn't concede that it's impossible, either). The point of generalism, and the point of a course on general CT, is for students to learn high-level principles of CT through explicit instruction *at this general level* along with practiced application of these principles with as diverse a set of cases as is feasible. (I'll return to this in the next section.) No generalist will dispute the importance of domain knowledge for deciding what to believe. For many problems, there is no substitute for knowing things, and general CT skills all on their own will be of limited use. That's a good reason for people to devote time to acquiring knowledge. It's not a good reason not to devote time to acquiring thinking skills to apply with and to that knowledge. The difficulty that people experience in transferring skills to novel domains, especially domains in which they have limited knowledge to draw on or to orient themselves, does explain some of the difficulty of teaching CT. But this appears to be a general obstacle in a lot of education, academic and non-academic, and there is no special reason to fear that CT is so vulnerable to it as to render a CT course futile.

Instruction in general CT skills abstracted from any meaningful content is not going to transform students into high-performing critical thinkers, especially if students are not also trying to learn about the world and engage with epistemic and practical problems. But it's reasonable to expect that acquiring and employing thinking skills even in somewhat superficial examples is a useful, perhaps necessary if not sufficient (Ikuenobe 2001), way to promote their deeper development in specific domains beyond those that can reasonably be visited in a CT course. The largest review (Abrami 2015) of attempts to do anything like this, a meta-analysis that included 684 published studies, finds considerable variation in the effectiveness of CT pedagogical design. The overall finding suggests a "modest but robust effect favoring the instructional viability hypothesis" and "the potential to teach generic CT skills" (2015: 282, 293). Under the right conditions, these skills are teachable and learnable.

The vocabulary of CT and the learning objectives of a CT course (this one or others with different emphases) constitute a "shared language" (Robinson 2011) with which to describe epistemic and practical successes and failures. Crucially, it's a shared language with which students can make sense of the epistemic and practical challenges they face and see others facing. It gives them a way to connect a problem in one domain to a problem they've seen in another, and to exploit lessons they may have learned about it already in that other domain rather than starting over, trying to build up new CT skills for every new domain. And it gives them a way not just to think about the problem itself, but to think about their own thinking about the problem, to monitor or reflect on their own use of CT techniques and devices (so-called *metacognition*[7]).

There is such a thing as general CT skills and there are good reasons to try to teach them.

1.3.2. Models of Critical Thinking Education

CT is not domain-specific. It is a genuine subject in its own right that can be meaningfully studied as general skills. Whether a student studies it directly or as an aspect of other subjects or both, it makes sense to expect them to become a better critical thinker in general. This conclusion from the previous section leaves many questions about effective and ideal CT pedagogy unanswered. It vindicates the model of teaching CT through a standalone course only in the weak sense that it answers a challenge that claims this model is a complete waste of time. It says nothing about how such a course should be designed to accomplish its goals.

An influential classification scheme of different models of CT education is from Ennis (2013). He contrasts the model under discussion here, the separate (standalone, dedicated) course, with an "embedding" model in which CT principles are integrated into courses on other subjects. It's expected (or hoped) that students have an intellectual investment in the subject of a course that they've chosen and are devoting significant time and energy to, and that investment can be harnessed for CT development.

Embedding may take the form of "immersion" or "infusion". Immersion is the embedding of CT skills and dispositions implicitly in the activities of the course, the idea being that students will absorb the CT as they work through those activities, work out their general, transferable form, and spontaneously deploy them in other contexts. Infusion is the embedding of CT skills and dispositions by explicitly introducing and highlighting them in or prior to the activities of the course. Students are to learn general principles of good reasoning while engaged in a particular domain, becoming aware of the fact that they are applying general principles to a specific topic so that they grasp how the principles would apply to other topics, as well. Finally there is a mixed approach in which the other models are combined for a more comprehensive CT education.

---

[7] I see metacognition as an aspect of CT proficiency but not as the essence of or key to it (cf. "Critical thinking is thinking about your thinking while you're thinking in order to make your thinking better" (Paul 1992: 7)). I'm not convinced that better critical thinkers are necessarily doing *more* metacognition, and they may well be doing less of it, in something like the way people do as they acquire expertise in a sport or musical instrument.

Unsurprisingly, this is Ennis' preferred model. Importantly, this concept of mixing the *pedagogical models* is not the same as, and does not imply, a "synthesis" of generalism and specifism (Tiruneh *et al*. 2016, 485),[8] which I take to be conflicting claims about the nature of CT.

If CT were domain-specific, only the immersion model would make sense, although even then, only when it's understood that students are learning to think critically about this or that particular domain, not in general, which specifism claims is an impossible or empty achievement. I've claimed that specifism is wrong, so the immersion model is not the only option, but that doesn't mean that immersion has no merits. Partly this is because specifists are right that the practice of thinking critically is thinking about some particular object of thought and one's ability to do that may depend greatly on one's knowledge of the relevant domain, especially where CT involves formulating predictions, considering alternatives, noticing background assumptions, estimating probabilities, etc. It's reasonable to expect that students will have, or be in the process of acquiring, deeper domain knowledge while taking a course in that domain.

There is some recent support (Tiruneh *et al*. 2016) for the efficacy of the immersion model in a physics curriculum. The study does not presuppose that CT is domain-specific, and tests only whether the course design improves domain-general skills such as reasoning, argument analysis, and hypothesis testing. Using a pre-test/post-test measure of CT and comparing to a control group of students in a regular physics course, Tiruneh *et al*. find a greater improvement in what they classify as domain-specific CT skills in a group of students who take a course designed "with an implicit focus on the desired CT outcomes as an integral part of the domain-specific classroom activities" (2016: 501). However these students show no greater improvement in what the authors classify as domain-general CT skills, and the study does not attempt to test CT skills transfer to other domains. Although it does not show that immersion is preferable to any other model, it does shows that immersion can work, at least by domain-specific success criteria.

The lack of evidence that teaching CT through the immersion model results in spontaneous generalization or transfer of CT skills accords with the expectations of most generalists. For example, Ennis is confident that an "explicit principles approach… providing something clear to students that they can remember" is the best approach (2013: 11). Tim van Gelder dismisses the

---

[8] Davies (2005) also frames his whole discussion this way, mistakenly construing the generalist-specifist debate as a "fallacy of false alternatives" (they are in fact alternatives), when his point is really the pedagogical claim that a "combinatory-'infusion' approach" is optimal. The success of CT infusion would not call for a concession by generalists or even a moderation of their position. It is fully consistent with, even predicted by, generalism.

notion that students will “become excellent critical thinkers merely by studying history, marketing, or nursing, even if their instruction is given ‘critical’ emphasis (as it should be). Critical thinking must be studied and practiced in its own right; it must be an explicit part of the curriculum” (2005: 43). Critics (Bailin *et al*. 1999b) of the domain-general view also endorse infusion. Even the specifist Willingham urges that “these [domain-specific] skills should be explicitly taught and practiced — there is evidence that simple exposure to this sort of work without explicit instruction is less effective” (2020: 44). It’s an exasperating position for him to take since his examples of “more specific skills” are things like considering the reliability of a source (domain: history) and understanding the relationship between a theory and a hypothesis (domain: science). If this is what he means by specific, it’s unclear what would count as general.

Willingham may have in mind something like Tricot and Sweller’s definition of a *general skill* as one that “can be used to solve any problem in any area” (2014: 266). Defined this way, as not only ranging over every subject but as useful for *every* problem, general CT skills merge into the concept of *g*, general cognitive ability, or intelligence, perhaps combined with a few basic dispositions.[9] As Tricot and Sweller note, such traits are so important that evolutionary considerations should lead us to expect them to develop through normal human experience and not be sensitive to anything as contingent (and modern) as formal education. And indeed research on intelligence and education does seem to show that cognitive abilities gained through education are “not mediated by *g*, but consist of direct effects on specific cognitive skills” (Ritchie *et al*. 2015). But generalists do not propose notions of CT so general that a course on these skills would amount to “trying to teach cognitive-linguistic IQ” (Scriven 2021: 34). A standalone CT course *does* aim to enhance “specific” cognitive skills, but not so specific that they can be learned and applied only as an aspect of some program of study whose primary focus is the knowledge and skills of a particular discipline.

Ennis’ (2013) ideal is a combination of separate dedicated CT course(s) and coordinated infusion throughout much of the curriculum at a fantasy institution he calls Wisdom University where the development of subject-specific CT and general, everyday CT is a core part of the institutional mission. Scriven’s somewhat less ambitious plea along these lines is for a kind of Wisdom minor

[9] Deductive logic comes to mind as potentially general in this sense, but while deductive logic must in some sense *apply* to every problem, this is not the same as being useful for solving every problem. Tricot and Sweller suggest as a general skill “thinking of similar problems with known solutions” and link general skills to our “biologically primary knowledge” that includes things like means-end reasoning, speech, and the ability to recognize faces (2014: 266).

program within existing universities, “a sequence of three or if necessary more standard courses” (2021: 34). Wisdom University has yet to be founded. And it’s a safe bet that colleges and universities will not be handing the reigns of academic programming over to philosophers anytime soon, notwithstanding the prominence of CT among institutional learning objectives. What exists are standalone CT courses that must do what they can without any practical way to coordinate their content with other courses or departments. The question is what lessons for designing a standalone CT course can be drawn from the discussion of CT generality and teaching models.

1.3.3. Lessons from the Generality Debate

Here are four lessons I draw for designing a CT course based on the preceding discussion of CT generality and pedagogical models.

1) Explain the nature of the course.

The first step in developing general thinking skills is to explicitly recognize that that is our goal. For many students who take this course, studying a subject of this degree of generality can be disorienting, since the topics of the examples and problems can seem to be one random thing after another. In its independence from, and applicability to, all manner of different topics, it is not altogether different from some language and mathematics courses, but the generality of these subjects has been recognized by most students since elementary school. Many, though they may have encountered the term *critical thinking* before, have little sense of what it means to study general thinking skills. I explain to students at the outset of the course and throughout that the objective is to gain skills that are not tied to specific topics or domains. Particularly at the beginning, this comment doesn’t mean much to some students, but for many it provides some useful orientation. The *tools* metaphor is helpful for this.

2) Make deep structure salient.

Even if his conclusions are unwarranted, there is something right about Wellingham’s (2010) analysis of the difficulty of teaching CT as a project that confronts students’ natural tendency to exhaust their cognitive resources on the surface structure of a problem. This leads them to see each problem as a novel challenge even when there is deep structure there to be exploited, and even when the skills of the current unit of study have been introduced and articulated in terms of that deep structure. I think the degree to which this phenomenon challenges students, and by extension their teachers, varies by student group and is not insurmountable. But the challenge is real and calls for direct engagement. Deep structure can refer to many aspects of a problem or its solution. One thing it refers to are the argument patterns and probability rules, which I highlight in the course manual with the consistent use of a simple bold frame.

For example, Units 4, 8, and 10 each present an argument pattern (IBE, analogical, and causal) that may be discerned in realistic examples of these types of argumentative writing, (acknowledging some implicit statements). The argument pattern is the deep structure of the examples that students see, and recognizing it is key both to the clarity of a reconstruction and the appropriateness of an evaluation. Students tend to get absorbed in the details of the example and fail to notice cues about the type of argument they're reading, which could prompt them to analyze it as an example of a general pattern or type of argument and to look for characteristic weaknesses. The key to the reconstruction is for students to work out what role each statement plays in this type of reasoning. In class and in practice exercises, I repeatedly draw students' attention back to the patterns, which many immediately set aside once given an example to analyze.

3) <u>Teach for transfer.</u>

The application of a general principle or pattern to specific concrete cases is one sense of *transfer*, but this term mostly refers to skill transfer from one concrete situation where it is learned to other novel situations. The difficulty of doing this, for many people, for many skills, is a point of agreement for generalists and specifists. Specifists draw a more radical and pessimistic conclusion, although some generalists share some of that pessimism. As van Gelder puts it, "the closest thing we have to a solution to the transfer problem is the recognition that there is a problem that must be confronted head-on… we must teach for transfer" (2005: 43). In choosing the examples to do this with, there is a trade-off to be considered. Examples that are similar in topic and structure can feel less disorienting to students; diverse examples can better serve to highlight the abstract principle or pattern being transferred.

Here's an example topic from the course (see Appendix E) in which I try to build understanding with similar cases and then deepen it with diverse ones. Unit 9 is about generalizing from a sample. It addresses (without formulas or calculations) the problems of insufficient sample size and selection bias. Along the way, I devote significant class time to the concepts of confidence and precision, which are routinely conflated. Grasping that these are distinct but related concepts enables a highly useful way to characterize a kind of epistemic limitation and trade-off that occurs in a wide array of situations. I use the standard type of case (surveys, etc.) to show how precision is specified at some (perhaps unstated) confidence level, and that an increase in precision (e.g. a smaller margin of error) comes at the cost of lower confidence, assuming the same sample size. Then I pivot to something that appears completely different, the ingenious though invariably misunderstood “forecast cone” of weather graphics, used to show the expected path of a major storm such as a hurricane. (Since graphics for the general public must be reliably self-explanatory, news agencies have been giving up on forecast cones despite their elegant informativeness.) This also communicates a statistical estimate that has a quantified precision, but it incorporates a time dimension. Some acceptably high (but lower than one might expect) confidence level is chosen and the available data is used to determine the precision, which naturally decreases into the future, hence the widening cone shape. The width of the cone at each time point corresponds to the error bar that students have by this point seen on a bar chart statistic. Similar cone-shaped representations of, for example, estimates of the human population over the coming century are less likely to be misinterpreted since, not being superimposed over a map, the cone does not present itself to the viewer as some kind of physical object. The two very different cases (poll and forecast cone) and the intermediate case (population projection) together highlight the underlying concepts.

The general skill here is applying this pair of concepts together to describe an important type of knowledge. It’s transferable even to cases in which the numbers are not based on measurements of samples and the mathematics of sampling doesn’t strictly apply. It enables people to understand what is being communicated in scientific estimates such as the temperature increase in a climate change scenario, the death toll in a COVID wave, or the age of some artifact based on radiocarbon dating. But it also gives people a way to think about how many guests they expect at a party, perhaps based on nothing more than Facebook invitation responses or something completely subjective like perceived enthusiasm. We’ll *probably* have *approximately N* guests: we’re 80% sure the number will be within some narrow range around *N* and 95% sure it will be within some larger range around *N*. Just the ability to formulate and communicate an estimate in this way, even if it’s not calculated, can help us think about how to plan for the party. In tutorials, students write and then informally explain their estimates, at multiple confidence levels, of things like the date that humans will land on Mars, the age they will die, the number of humans who have ever lived, and so on. These are of course not confidence intervals in the mathematical sense since they express only subjective probabilities and are not calculated with the tools of statistics, but not a small fraction of students finally get the notion of a trade-off between confidence and precision in this little exercise, and sometimes offer interesting and thoughtful justifications for estimate ranges that depart from the general pattern of decreasing precision with increasing confidence, such as some personal reason for a hard upper limit (unchanged even at higher confidence) on the estimate range for their age at death.[10]

This non-mathematical but also non-superficial treatment of “probably approximately correct” thinking is an example of how I try to identify and present a profitably transferable general CT skill. It is not bound to any particular example or scenario. Students can learn to recognize and informally use it with diverse examples, and transfer of the analytical and terminological skills to the new examples deepens their new knowledge rather than becoming awkward, forced, or simply unhelpful mental gymnastics. But it’s certainly not such a general skill that “teaching” it amounts to a mere slogan or to the likely futile task of trying to boost general intelligence.

4) Internal infusion.

In the infusion model, CT is taught as an aspect of courses on various subjects in order to draw on students’ deeper knowledge of and greater engagement with these subjects. By *internal infusion* I mean a course design that attempts to capture the spirit of infusion, and some of what makes it a good teaching model, but within the confines of the standalone course in which CT itself is the topic of study. It’s *internal* in contrast to the mixed model where the infusion component is external to one or more courses dedicated to general CT. The best way to do this is with an assignment, although this requires careful time management, especially when the course uses TAs whose available hours are strictly limited by their contracts. The hypothesis testing assignment based on Units 4 and 7 (see Appendix B) is an attempt at internal infusion. Students choose any topic that interests them, provided that some recently published news item supplies an observation or evidence detail. The assignment is also an opportunity to incorporate two of the elements of CT instruction that Abrami *et al*. (2015) find in their meta-review to be most effective: authentic/anchored instruction and mentoring. By authentic/anchored instruction, they mean “a well-defined real-world problem that the students are analyzing”. It includes applied problem

[10] There are many other issues to confront in helping people think about quantitative estimates. For example, Michael Smithson (2016: 484) discusses evidence that people prefer greater “informativeness” (precision) not only when they understand that a more precise estimate entails a lower *chance* of accuracy (of including the real value) but even when they are told that the more precise estimate is *in fact* inaccurate (fails to include the real value).

solving and case studies among other formats. Mentoring is normally one-on-one teacher-student interaction.

The assignment involves formulating alternative hypotheses, along with informal justifications of the assigned prior probabilities, and clearly describing some actual or hypothetical additional evidence, along with informal justification of an estimate of evidence strength. Applications of Bayes' rule show the updates based on possible evidence discoveries. Although it's not an essay, the description and justification sections require some thoughtful writing, and the exercise is not a fictional scenario as are many of the practice questions. Students engage with a real-world problem of belief updating, which they have discovered themselves, in some depth by drawing on new and background knowledge.

The assignment normally incorporates a time-intensive mentoring stage in which every student meets privately with me to discuss their topic and some portion of work in progress, minimally their alternative hypotheses and additional evidence. In this I can provide encouragement and show interest in the topic they've chosen, check that they've formulated a clear question (typically a call for an explanation), and help them ensure that their hypotheses are suitably formulated to answer their question. This process also makes the assignment about as resistant to plagiarism and contract cheating as any work completed out of class can be.

Part 2. Course Design: Parameters

Introduction

Part 2 describes what I call course *parameters*. These are not topics or units or learning objectives but rather design aspects of a CT course, ways the course is structured and presented, and ways that the highly limited resources of class time, student focus, and available instructor/TA assessment hours are allocated. The five parameters are labeled as oppositions but only parameters 1 and 5 are direct oppositions. The oppositions are partly products of the need to choose what to emphasize in a single-term, introductory level course.

The teaching resources are two tightly coordinated components: a textbook (which I call a course manual – Appendix C) and a set of highly crafted PowerPoint slides that structure the class activities. The manual can be read on its own, whereas the slides are purely classroom resources to discuss and to prompt student engagement. The slides do not summarize the manual. Their extensive animations and lack of explanatory text require close familiarity for effective delivery, but TAs given select slides to prepare tutorials can quickly learn to use them effectively. The course is designed to be ideally delivered with a four-hours-per-week tutorial-style class, although it can be adapted to three hours or two hours lecture plus one hour tutorial. The parameters of Part 2 mostly describe the two components together, which is to say, they mostly describe the content itself. Parameter 5, specifically pertaining to the course manual, is the exception.

2.1. Parameters

2.1.1. Unified vs. Pluralist Approach to Reasoning

Parameter 1 is the extent to which the course's treatment of reasoning is based on one principle or tool, such that most of the major topics and problems are framed as opportunities to apply this general reasoning principle or tool. In this course I take an approach I'll call restrained pluralism.

Unified here does not mean forcing every single task in the course into the exact same format, a sort of thinking hammer for which every case becomes a nail. But it does mean that students learn one fundamental way of performing or representing a reasoning task, and they learn nothing else as an alternative way to think about what to believe and do. I see at least two plausible candidates for such a unified approach. One candidate is the reasoning device of an argument, perhaps even specifically a deductive argument in the spirit of Whately; another is a Bayesian updating procedure. A third possible candidate is the cognitive psychological view of mind as a marvelous but flawed inference engine that employs various heuristics and is vulnerable to numerous cognitive biases that may be counteracted but not fully suppressed — a "debiasing" approach (Kenyon and Beaulac 2014).

A fully unified approach is an extreme design decision. Probably no or few CT courses are fully unified, but some are more unified than others. For example, there is the Woods, Irvine, and Walter forthrightly titled CT textbook *Argument* (2000). In Manley's *Reason Better* (2019), he criticizes textbooks that "introduce numerous 'forms of reasoning' as though they were a grab bag of techniques that have nothing to do with each other", favouring instead a Bayesian "unified account of evidence". A unified approach has some compelling points in its favour. One is organizational economy. A student in their course review once complained that the (non-unified) course has "so many theories". This particular student engaged only superficially and intermittently with the course, but I want to make sure that more serious students don't also have this experience. In a course where some students feel overwhelmed by the (unfortunately) new demands of careful analytic thought and language, an approach in which they could gain confidence with a single widely applicable reasoning tool has some merit.

But this is not to say that it's the on-balance best approach. The unified approach encompassing at least all the non-deductive (IBE, analogical, generalizing/statistical, causal) inferences such as Manley (2019) builds on the basic fact that for any premises of any form to raise the probability of a conclusion, they must be more likely supposing the conclusion true than they are supposing the conclusion false. This technically unifies all these inferences into a Bayesian scheme. For example, one can explain why generalizing from a larger sample is better than generalizing from a smaller sample by noting that the sample measurement is more likely to be what it is if it matches the population, and less likely to be what it is if it doesn't match the population, when the sample is larger compared to when it is smaller. The question is whether this enables a significantly more coherent treatment of these forms of inference that makes learning easier or deeper. I include some elements of this type of analysis in class. I point out roughly this explanation for the relative value of larger samples before showing how this value gets cashed out in the form of improved precision, and I point out that the structure of a controlled experiment (measuring the dependent variable in an experimental group and a control group) can be seen as a kind of Bayesian process. But these are not major lightbulb moments for most students that suddenly show them the essence of these forms of reasoning. I see no indication that investing more heavily into presenting these inferences as, at bottom, the same thing is going to help students do what I want them to do, even if the Bayesian characterization is ultimately correct, as I believe it is. People appeal to explanatory goodness and analogies, they generalize from limited data, and they base claims of causal efficacy on observations that some factor is the only difference when some effect occurs; the overall project in these units is to equip students to comprehend, evaluate, and communicate about these types of reasoning in the terms in which they will encounter them. If comprehension and evaluation were inevitably superficial, or evaluative criteria had to be just stipulated as arbitrary tests, without an explicitly Bayesian interpretation, the case for a unified treatment would be more compelling. But this isn't the case. Just because it is *true*, if the discovery of a lone difference supports a causal claim, that the lone difference is more likely given the truth of the causal claim than it is given its falsity, doesn't mean that this fact is the only proper way to understand the support conferred by the discovery on the causal claim.

There is no defined extreme that is opposite to a unified approach, since there is no real limit on how disunified a course might be, but the contrast is a pluralist approach where students learn different reasoning tools for different problems or even for the same ones, and no reasoning device is presented as dominant or fundamental. A pluralist approach gives students a variety of tools and shows them how to use the "best one for the job". At least as important is the value of a pluralist approach for equipping students to comprehend the different forms of evidence-based persuasion they encounter. To the extent that a CT course should aim to do this, the case is stronger for a pluralist approach. This course is committed to that goal.

The more abstractly one defines an approach to reasoning, the easier it is to see a course that uses that approach as highly unified. For example, Haber suggests "structured thinking" as a core component of CT, more expansive than logic, that better captures the way in which CT is about "disciplining ourselves to think in an organized fashion" (2020: 38). At this level of abstraction, this course is quite highly unified since I emphasize structured thinking throughout. Nearly every unit features one or a few basic structures of reasoning (deductive and inductive argument patterns, or probability rules) that underlie the examples and problems, and which can or must be reconstructed or constructed in order to accomplish important objectives of the unit. However I think of unified and pluralist approaches in terms of skills. I don't see structured thinking as an approach to reasoning in this sense. It's closer to a disposition, a habit of seeking or creating order in the interest of productive reasoning and communication.

Other than structured thinking, the most obvious sense in which this course is unified is that arguments are a thread running through it, from Unit 1, a brief overview of the concept of an argument and the main elements of argument analysis, to Unit 10, which introduces the last of the course's argument patterns. But arguments are not an obsessive focus, as there are major sections on other topics in Units 2 (counterexamples), 5 (confirmation bias), 6-7 (probability and Bayesian reasoning), and 10 (controlled experiments). The recurring appearance of arguments, when they are useful, reinforces that the ability to think about arguments is indeed central to CT. But there are many other things to accomplish.

Compare this to the treatment of arguments in the popular textbook *The Power of Critical Thinking* (MacDonald and Vaughn 2019). Arguments are introduced in one section of Part 1 (of 4) on "basics". Part 2 reminds readers of arguments in a couple of sentences, then sets them aside to look at good and bad "reasons" for belief, making almost no use of the argument as a format for organizing or presenting reasons (premises) for belief. Then Part 3 on "arguments" has two sections on deductive arguments and one on inductive arguments. Having finished with arguments, the authors immediately turn, in Part 4 on "explanations", to inference to the best explanation, as though this were not a type of argument.[11] The problem here is that MacDonald and Vaughn treat arguments as a topic rather than a tool. Whereas I introduce the tool and then use it as appropriate, they try to gather up the argument elements of CT for one part of their course. But this just doesn't work, and they have to sneak in arguments elsewhere or awkwardly talk about how to argue without talking about arguments.

At the same time, it would be a distorted view of CT, and probably a dull experience for students, for the course to be devoted entirely to arguments. Hence the diversity I summarize above. But this diversity is not simply a matter of doing things other than arguments. Since I teach arguments as a tool not a topic, it's natural to look at some topics with a variety of tools, including arguments. The Unit 4 treatment of IBE and confirmation/disconfirmation and the Unit 7 treatment of Bayesian updating jointly do this for scientific reasoning. I'll return to this example in section 3.2.

The other main example of this sort of dual treatment is in Unit 10 on causal reasoning (also a type of scientific reasoning, to be sure, but positioned to build on the statistical reasoning of Unit

[11] Battersby (2006) reasonably suggests that IBE would be better named "argument to the best explanation".

9). I believe the most common and useful type of causal reasoning is Mill's method of difference.[12] It's presented as an extended argument with a structure similar to the IBE and analogical arguments in earlier units. Students see partial and complete real-life examples that build up to the full extended pattern, with which they reconstruct arguments and articulate a criticism of their characteristic flaw, the failure to account for other relevant differences that are potential causes. When we turn to controlled experiments, this problem recurs as the problem of confounding factors, whose potential causal roles may be depicted in simple, intuitive arrow diagrams depicting common causes. A strong sub-argument (supporting a "no other relevant differences" intermediate conclusion) corresponds to an experiment in which variables are well controlled, either directly manually or indirectly through a randomization procedure. This pluralist treatment of a single topic demands extra time compared to a unified approach but the trade-off is worthwhile. Students gain a deeper grasp of the concepts and a better ability to recognize and discuss the causal reasoning they may encounter both in informal, everyday contexts and in more formal contexts such as primary or secondary social science literature. Along with the diversity of topics which make the course much more than a tour of arguments, the dual treatments of some major topics justify calling the course's approach to reasoning pluralist. But it's a restrained pluralism because I take Manley's "grab-bag" point to heart and have designed the course to present a limited set of thinking tools, each employed as appropriate for topics arranged in a natural progression.

A more unified approach to non-deductive (IBE, analogical, generalizing/statistical, causal) inferences such as Manley (2019) builds on the basic fact that for any premises of any form to raise the probability of a conclusion, they must be more likely supposing the conclusion true than they are supposing the conclusion false. This technically unifies all these inferences into a Bayesian scheme. For example, one can explain why generalizing from a larger sample is better than generalizing from a smaller sample by noting that the sample measurement is more likely to be what it is if it matches the population, and less likely to what it is if it doesn't match the population, when the sample is larger compared to when it is smaller. The question is whether this enables a significantly more coherent treatment of these forms of inference that makes learning easier or deeper. I include some elements of this type of analysis in class. I point out

[12] I present the method of difference but not the method of agreement as the main pattern of everyday causal reasoning. In most real examples of causal reasoning, even when agreement language is present, it is actual or presumed difference (or correlation) that supports the causal inference, and the method of difference roughly corresponds to the experimental procedure (controlled experiments) that is the focus of the unit.

roughly this explanation for the relative value of large samples before showing how this value gets cashed out in the form of improved precision, and I point out that the structure of a controlled experiment (measuring the dependent variable in an experimental group and a control group) can be seen as a kind of Bayesian process. But these are not major lightbulb moments for most students that suddenly show them the essence of these forms of reasoning. I see no indication that investing more heavily into presenting these inferences as, at bottom, the same thing is going to help students do what I want them to do, even if the Bayesian characterization is ultimately correct, as I believe it is. People appeal to explanatory goodness and analogies, they generalize from limited data, and they base claims of causal efficacy on observations that some factor is the only difference when some effect occurs; the overall project in these units is to equip students to comprehend, evaluate, and communicate about these types of reasoning in the terms in which they will encounter them. If comprehension and evaluation were inevitably superficial, or evaluative criteria had to be just stipulated as arbitrary tests, without an explicitly Bayesian interpretation, the case for pluralism would be weaker. But this isn't the case. Just because it is *true* that a lone difference is more likely given the truth of the causal claim than it is given its falsity doesn't mean that this fact is the only proper way to understand how it supports the causal claim.

2.1.2. Skills vs. Dispositions

Parameter 2 is the relative emphasis on teaching CT skills vs. instilling or inspiring CT dispositions. This parameter is set most of the way to a skills emphasis in that the topics, activities, and assessments are directly focused on skills, though their presentation at least models some important CT dispositions, and the assignments and class discussion give students some opportunity to develop these. In this section I'll briefly explain why this course is like this. Although this section doesn't attempt to justify teaching CT in a dedicated, standalone course (as I addressed in Sec. 1.3), it considers an implication of doing this, and so addresses the first question in the introduction, about how the course is designed to make the most of its limited opportunity to shape and promote the set of traits that is CT.

It would be an unusual CT course that focused directly on CT dispositions with "disposition" units or exercises included at the expense of "skill" units or exercises. A skills-based CT course is the norm. Of course that is also hardly unusual among academic subjects. In mathematics, chemistry, history, or psychology, etc., shaping the character or personality of students is not typically understood to be a pedagogical objective. The successful student is one who demonstrates mastery of the subject in the relevant academic or professional setting, not someone who becomes a certain kind of person by endorsing and acting from certain values. But CT is different from these other academic subjects in that dispositions appear prominently in most influential definitions of CT, as well as my own. A critical thinker is more than just someone who does certain things in the relevant academic or professional settings, but someone who internalizes certain motivations, habits, traits, and intellectual virtues, and exhibits these dispositions in their life generally. An extreme view would be an entire dispositional theory of "good thinking", including what most would recognize as CT (Perkins *et al*. 1993), with ability understood as a component of disposition, along with inclination and sensitivity to appropriate context. A more standard view is that abilities and dispositions are separate components of CT, with neither a component of the other. This is the view stated directly in Glaser's (1941) dispositions-knowledge-skills definition.

Glaser himself appears optimistic about the prospects for people learning CT dispositions: "The aspect of critical thinking which appears most susceptible to general improvement is the attitude of being disposed to consider in a thoughtful way the problems and subjects that come within the range of one's experience. An attitude of wanting evidence for beliefs is more subject to general transfer" (1941: 175). But Glaser's optimism is tested by ubiquitous and striking cases of failure to transfer this attitude. People are adept at compartmentalizing their feelings of the need for evidence. One thinks of the person willing to believe in angels or the QAnon conspiracy theory or the professions of love from an internet romance scammer, based on the flimsiest evidence, while demanding absolute perfection in the evidence for climate change or evolution or vaccine safety. Moreover transfer and improvement are not exactly the same thing, even if we'd be more likely to say that someone's attitude of wanting evidence improved if we see them transferring that attitude to novel scenarios. About improvement, there are several distinct questions: 1) Can CT dispositions improve at all or are they fixed character traits? 2) Can they be improved by formal education? 3) Can they be improved by a standalone CT course? Optimism about both (1) and (2) is necessary but not sufficient for optimism about (3). And then, even if a standalone CT course is a mechanism by which CT dispositions could be instilled, one confronts two further questions: 4) Is there any practical way to reliably assess students' dispositions within a CT course? 5) Is it even legitimate to assess a student based on the kind of person they are and the values they hold?

David Manley's *Reason Better* (2019) illustrates the pedagogical problems of how to instill and assess dispositions. Manley sees curiosity as the first key attribute of the CT "mindset" (the others being thoroughness and openness) (Ch.2 [ebook]). This is a rousing claim that identifies what surely is an important aspect of CT, which I take to be roughly equivalent to caring about truth and evidence, as in my account of the successful critical thinker. But here is Manley's immediate explanation (definition?) of curiosity: "The goal is for our beliefs to reflect how things really are; this is best achieved when our confidence in a belief matches the strength of the evidence we have for it." How does this explain the concept or application or virtue of curiosity? These are literally the first words he writes after "1. *Curiosity*". Immediately he shifts to something clearly distinct from curiosity, and more testable. Then, adapting Julia Galef's (2021) vivid distinction between the "soldier" (defending a position) and "scout" (looking for truth) mindsets, he urges students to "learn how to feel differently" and arrives at this: "Genuine curiosity matters more than any specific reasoning skills we can acquire from academic training."

This is all fine to say to a person who is antecedently interested in (and so likely disposed toward) CT. But the large majority of students in this course are not reading the textbook for exhortations like this. (I'll return to this issue in Parameter 5.) They are reading it (if they are reading it) looking for something concrete to learn and do. And what is that, in this case? We see this, presumably, when we see how Manley envisions assessment of the "Mindset" chapter. One implied way to assess this appears in the following pages, not as a way for an instructor to determine if students are themselves becoming more curious, but rather as an alleged way for students to diagnose insufficient curiosity in others: basically, they commit the red herring fallacy or exhibit wishful thinking, conventional CT concepts, though Manley doesn't use either term, instead introducing his own seemingly redundant term, a "suspiciously unexpected set" of beliefs. The banks of sample exam questions that show how Manley envisions assessment of this chapter are silent on the dispositional elements like curiosity. The closest these come to trying to assess curiosity are some multiple-choice questions about, for example, the value of thinking with degrees of confidence, which Manley has somewhat oddly directly linked to curiosity. Of course, not every single word of a textbook has to be directly testable, especially if the words help explain something else, such as the cognitive biases that get significant attention from Manley. But that's not really the case with curiosity. It's possible but of dubious helpfulness to cast errors such as confirmation bias and the availability heuristic as species of the personal character flaw of "lack of curiosity". This example of a core CT disposition given a prominent place in a standalone CT course, then "taught" by exhortation and not meaningfully tested, gives some indication of why this course has a very light emphasis on CT dispositions.

There are more and less comprehensive visions of the intellectual virtues of the critical thinker. A minimal vision — roughly the one I incorporate into my definition — is someone who values truth and evidence and strives to follow standards of reasoning that make the best use of their evidence. I previously mentioned the more comprehensive vision of the California Foundation for Critical Thinking, which includes traits such as humility, autonomy, empathy, and integrity. I'm agnostic about whether anything like this more comprehensive set of dispositions can be instilled through higher education, and I highly doubt that it can be instilled through a single lower-level course. But while it makes little sense to try to directly teach an attitude of caring about evidence or a motivation to reason in ways that make the best use of it, the activities (and instructor) of a standalone course can at least model these attitudes for students and affirm the intrinsic and instrumental value of truth and reason. And a skill-focused course can give students opportunities to reflect on these values and discover their own identification with them as they apply CT skills with topics of their own choosing. This is another reason for the internal infusion aspect of the course design, and another reason why, even in a quantitative course with marking hours limited by TA contracts, I incorporate some activities directed by students' own interests. Of course it's also another way in which one would expect the best results with the full-scale CT infusion of Wisdom University.

The fact that CT, probably uniquely among academic subjects, has values and character traits as part of its definition does not mean that it's practical or justifiable to evaluate students on these things, even if it is possible to promote some core CT dispositions. This would require me or a TA to know each student in a personal way that goes beyond what is possible through exams and short assignments and even brief one-on-one interviews. And it would not be justifiable within an educational framework committed to liberalism, even if CT is one of the highest ideals of liberal education. We can't assign grades to students in CT based on what kind of values they appear to hold any more than we can in Mathematics or English or Business or Fine Arts courses. It's just an awkward fact about CT that, unlike these other subjects, there is an entire dimension of our conception of the subject that is more or less off-limits for evaluation, at least by the standard mechanism of assigning grades (other assessments such as letters of reference in special cases might be a different matter). What we can practically and justifiably evaluate are students' demonstrated abilities. So the development and assessment of CT skills are the direct focus of this course.

Fortunately CT skills education is a rich project of considerable value. The fact that the sort of morals education suggested by comprehensive definitions of CT is mostly not viable in a standalone course remains an awkward feature of CT education. What I can do is empower students to apply some CT skills to some of the important information, problems, and tasks they may encounter. I can model through examples — whose first function remains to make difficult and new ideas clear — the value of reason and evidence, as well as epistemic humility, interpretative charity, and the other intellectual virtues. And I can encourage students through discussion and small assignments to discover, develop, and deepen these values in themselves.

2.1.3. Critical Thinking for Comprehension vs. Communication

Parameter 3 is the extent to which the course equips students to comprehend vs. communicate argumentative, explanatory, probabilistic, and scientific information. While the former might be construed as passive and the latter as active skills, I see this contrast more in terms of the direction of information flow (from world to thinker or from thinker to world). I think of comprehension as an active process that involves recognizing and categorizing information, noticing assumptions and relationships of support, and organizing information for more flexible reference. This is prior to appraisal, and prior to decisions about how to act on the basis of what is comprehended. Communication skills include the planning as well as the expression of reasoning and criticism.

This section loosely addresses the second question in the introduction, how the course serves students with a wide range of linguistic and educational backgrounds. An ability to function and participate effectively in the language environment is the foundation of a liberal education, and a CT course can play a valuable and distinct role in developing this ability. Of course some students arrive at higher education much further along in these abilities than others. I don't suggest that Parameter 3 is tuned precisely to help the weaker students, but its balanced setting in this course does reflect the importance of both types of skills and for this role of the course in students' education. A balanced setting isn't a radical choice, but I think it's worth discussing what that balance looks like for aspects of this course that are somewhat less prominent than are Parameters 2 and 4 in everyday references to CT such as those quoted in the introduction.

Some of the purest tests of comprehension in the course are the argument reconstructions in Units 4, 8, and 10. These require students to work through passages of argumentative text to identify and extract extended arguments (a main argument with a premise supported by other premises in a sub-argument), including implicit statements and alternative hypotheses, by discerning the underlying argument pattern or schema. The reconstructions use statement numbers from the paragraph, to be arranged in standard form (conclusion below a line and premises above and the intermediate conclusion in a box for emphasis) in a standardized answer format:

Main Argument	Sub-argument
⑤	③
⑥	④
———	———
②	⑥

I settled on this method of laying out an argument after experiments with argument mapping (connecting statements by arrows of inference) led me to conclude that it offers no additional benefits, at least for the paragraph-length extended arguments that are sufficiently challenging for a single exam question. Argument mapping is thought to enhance comprehension by "dual [multi-sensory] coding" and by relieving the burden on limited cognitive capacity by "'off-loading' information as visual display" (Davies *et al*. 2021: 118). As in argument mapping, in these reconstructions, students build up the argument by physically arranging statements on the page to indicate their relationships to one another. The same cognitive rationale applies. Like argument mapping, they function as both an aid to, and a demonstration of, comprehension.

These exercises also serve another important function for the significant number of students who arrive in the course with weaker reading skills, not immediately prepared to grasp the idea of a recognizable type of inference with characteristic liabilities. Many students have been trained, and in some cases continue to be trained, to read a text by searching for the "main idea" in each section or paragraph without any thought to inferential connections. And to be fair, many texts are written in such a way that this is not a completely misguided approach. Even in argumentative texts, many sections and paragraphs are not complete structured arguments. Students instinctively revert to main idea hunting when confronted with a complex passage, even one explicitly presented as an argument. While there is some loose sense in which an argument's conclusion might be thought of as the main idea of an argumentative passage, trying to comprehend an argument by seeing its conclusion is the most important statement and its premises as less important is hopeless. The demands of argument reconstruction force students to go beyond this, a key step in enabling them to read critically by apprehending the rationally persuasive content of a text.

Most of the course is designed to help students become informed citizens who can constructively participate in society's important discussions. In particular, I take the university's Science designation for this course as a serious obligation. Giving non-science students the competence, confidence, and hopefully interest to read mainstream science journalism is one of the most important aspirations of the course. Appendix A discusses two examples of this sort of writing along with explanations of how content from throughout the course's ten units prepares a student to engage with them. It's worth noting that if specifists were right that meaningful and useful CT is inherently bound to specific content knowledge, a course on general CT skills would have little to offer in these cases. But these examples show that the general skills covered in this course would significantly aid someone to grasp pieces of writing like these.

The communication element of the course includes some basic decisions such as my determination not to give in to the temptation to use multi-choice exams. There may be many reasons not to use multiple-choice exams for CT if it can be helped. Willingham notes that students focus on multiple-choice question answering strategies rather than CT and concludes that "assessing critical thinking requires that students answer open-form questions, and that means humans must score the response" (2020: 45).[13] Kathrin Stanger-Hall (2012) studied test formats in a biology course and concludes that the multiple-choice exam format encourages less "cognitively active study behaviours", stifles a variety of critical (higher-level) thinking skills and habits, and leads to worse outcomes. In this course, although some parts of some exam questions are effectively multiple-choice simply because they have a small number of possible answers, most questions use some kind of constructed response format that requires students to formulate and express a thought, that is, to actually communicate their thinking through the exam rather than merely select a presented option. Students must think of alternative explanations and background assumptions and convey them with sufficient clarity (Unit 4), explain specific forms and examples of confirmation bias (Unit 5), and organize and express evaluative responses to inductive arguments, including some with a normative character (Units 8-10).

The assignments on extended argument construction and hypothesis testing (see Appendix B) are opportunities beyond the exams for students to develop and exercise communicative skills, including a detailed planning of their written expression. With argument construction, this is a separate standard-form stage of the composition, preceding the written paragraph, which forces students to work out the underlying pattern of reasoning they will convey in conventional paragraph form. With hypothesis testing, students must carefully compose their explanatory and predictive statements and relevant background discussion in order to connect the written presentation with the quantitative elements of Bayesian reasoning.

---

[13] That's what it means as of 2023 February, when generative AI writing technology is rapidly advancing into mainstream applications and commercialization. How long this will be the case remains to be seen.

2.1.4. Reactive vs. Constructive Critical Thinking

For Parameter 4, I borrow the terms *reactive* and *constructive* CT from David Hitchcock, who defines reactive CT as thinking that "appraises others' statements and arguments" and constructive CT as "solving unstructured problems and making complex decisions" (2017: 492). In Part 1, I noted that Thayer-Bacon (1998) calls for a transformation of CT into what she calls "constructive thinking", which has different psychological and theoretical underpinnings. Hitchcock's concept is not so revolutionary. This is constructive thinking within CT, not instead of CT. With this parameter, the course's emphasis is set fairly far toward reactive CT. As with the balanced setting of Parameter 3, the reactive emphasis in this course is not in itself an unconventional decision. Hitchcock says that he taught entirely reactive CT for many years before expanding his teaching to include some constructive CT. Despite its reactive focus, this course does diverge from a conventionally reactive formula in some important ways, and it has some important learning objectives that are not well characterized using the reactive-constructive opposition.

A natural first reaction to Hitchcock's definition of *constructive* might be: "Yes, that sounds great! We need more of that in our citizens, voters, leaders, consumers, employees, etc." Reactive CT might be construed as lower-level skills, not unimportant but also not the pinnacle of CT. We see something like this way of thinking in the 2001 revision of Bloom's taxonomy, which re-orders the top tiers to demote evaluative thinking, from [1) Evaluation, 2) Synthesis] to [1) Create, 2) Evaluate]. But it would be a mistake to view a reactive orientation in a standalone CT course as setting low expectations for students. Construction and creation are valuable; appraisal and evaluation are valuable. Students taking this course are entering adulthood with the cacophony of the claims and arguments on the internet in their pockets. Having skills to impose some order on the parts of this they attend to, and to formulate thoughts and speech in response, is not a CT project of secondary or lower importance. The number of these claims and arguments that are of a broadly scientific character, and the prominence of analogical reasoning in ethical, social, and political contexts, justifies the attention I devote to reactive CT with these topics.

Reactive skills are also skills of comprehension. They can help someone who might otherwise regard themselves as a mere spectator to a discussion, who perhaps feels no need to form any opinion on it, to understand how someone else appraises an argument. This has the additional benefit of enabling them to form a useful opinion of the appraiser, whom they may have occasions to consult or trust in the future. And of course reactive skills are argument skills. When someone better anticipates what potentially critical appraisal their own view may encounter, they better understand whether their view is well supported and whether they ought to hold it. It's tempting to suggest that in this sense, reactive skills can be constructive. Hitchcock's definitions leave the place of argument construction unclear: Is argument construction constructive? The argument construction assignment is at least semi-constructive. It doesn't fully meet the definition because it's not a fully "unstructured problem", but even a short composition that must achieve logical, argumentative, informative, and stylistic goals could be thought of as a "complex decision".

The idea that people can learn good reasoning by learning about bad reasoning has a long history. Aristotle's *Sophistical Refutations* is not addressed to aspiring sophists but to aspiring philosophers. In *The Port-Royal Logic*, Arnauld and Nicole note that "examples of mistakes to be avoided are often more striking than the examples to be imitated" (1996: 189). I see little reason to think that mistakes, including fallacies, are usually more "striking" than examples of good reason, though I agree some may be. Regardless, I think that examples of bad reasoning are a valuable tool for teaching good reasoning. In some cases, their strikingness may be a factor in this, but the general reason is that they present students with a contrast, making the important features of the good reasoning salient. So, for example, when I first introduce students to specific patterns of good reasoning with valid syllogisms in Units 2-3, I place these directly beside the most common invalid patterns to make the essential features of the valid patterns salient. In Units 4-5, I clarify what good confirmation reasoning would be using several forms of confirmation bias.

As much as possible, students should learn good reasoning by learning what *makes* the reasoning good, not just by imitating it and not just by avoiding errors. For example, with the Unit 3 syllogisms, I spend significant time in class working with the concepts of sufficiency and necessity and the language of conditionals, pointing out not just their core meaning but the way they function in English, such as the way the …*only if*… construction typically stresses the necessary condition (the consequent). In this course, the validity of *modus ponens* and *modus tollens* is grounded simply in the meanings of these common conditional sentences. But students learn them better by noticing the contrast to the invalid Affirming the Consequent and Denying the Antecedent forms, whose invalidity is based on what their conditional premise does *not* say. It would be foolish to think, and I'm not aware of anyone who does think, that we can effectively teach good reasoning simply by teaching students what errors to avoid and to criticize in others. Hitchcock mocks this idea by imagining it in a different context, teaching tennis by teaching the tennis mistakes to avoid (2017: 405). But it's important not to lose sight of the legitimate role of these fundamentally reactive appraisal skills not just in criticism but in learning. Indeed, they play an important role even in tennis instruction, where the value of both descriptive ("you made this error") and prescriptive ("correct the error this way") feedback is well established (Reid et al. 2007: 8).

As I discussed in Part 1, an education well designed to promote CT is one thing and a well designed standalone CT course is another. My concern is the good design of a course with numerous educational functions (in particular the institutional Quantitative, Science, and Social Science designations) and the constraints that come with any course that has to be capable of being implemented in a large lecture format. The unstructured problems and complex decision-making of constructive CT are best suited to an infusion delivery model, where the focus on a particular topic allows the design of assignments and exam questions that draw on deeper knowledge of and engagement with a topic. To create a course that first-year students of diverse educational backgrounds can manage, that covers the logical, writing, argumentative, scientific, and quantitative topics that it promises to the university and to the Philosophy Department, the emphasis has to be on the more structured tasks characteristic of reactive CT. The quantitative topics, mostly in Units 6-7 on thinking with and about probability, aren't well placed in that opposition. These skills are neither reactive nor constructive in the strict senses defined above, though they can be put to good use in either type of CT.

2.1.5. Concise vs. Complex Writing Style

Unlike the previous parameters on the CT content itself, Parameter 5 concerns the primary vehicle for delivering the content, the textbook, which I call a course manual.[14] This section addresses the second of the five project questions in the introduction, about how the course serves students from a wide range of linguistic and educational backgrounds, which means how it reaches and engages students who are ill-equipped for post-secondary education while still challenging those who are prepared. The writing style of the course manual can be described with a parameter that takes values ranging from extreme concision to the extended prose of a book that takes the eager student on a reading journey. This parameter is set at extreme concision. The writing is as plain and direct in its vocabulary and sentence structure and as it can be. Although it is very much meant to be *read* — it's not a manual in the sense that it invites students to use it simply to look

[14] The secondary vehicle is the closely coordinated set of class slides, a few of which are demonstrated in the video appendices. They make no attempt to summarize the course manual because the manual is so concise that it cannot really be summarized.

up definitions or instructions as needed — my compositional goal for each unit is to present the least daunting reading task possible, with every section of every unit straight to the point, and every point directly connected to something they need to be able to do.

The decision is informed by my experience in university and college classrooms in the lower mainland of British Columbia, Canada,[15] with students who differ widely in their academic preparation, aptitude, and ambition, and in their language skills (and first languages). Some excellent students enroll in the course out of genuine interest and desire to improve their mind; many take it as a program requirement or an elective that fits their schedule. Early experiences teaching CT with some of the available published textbooks, and with early and differently written versions of some of my own material, led me to the conclusion that using these texts as documents for students to read, as opposed to documents for *me* to read and then convey to students what they "need to know", was aspirational to the point of being delusional. For all but a tiny elite group of students, these texts simply did not function as learning resources. The style and sheer amount of writing they contained, only some of which I could draw on for the course, was clearly a major factor in this.

The writing style of the course manual confronts the reality of students' highly variable ability and willingness to read, at least at the first-year level in the institutional context of its creation. This document attempts to strike a bargain with students: If they will put time into reading for this course, I promise them that their efforts will be unusually highly efficient. After teaching approximately 12 sections of the course at SFU, most with around 250 students, and far more sections at colleges, I have had only a couple of students express to me that they wished the manual were less blunt or that it elaborated or chatted more, but many express their appreciation for its extreme succinctness. This includes some students who evidently would not normally read textbooks at all for their courses but who in this case can be coaxed into giving the sparsely typed pages of this document a try. It also includes students who are evidently very capable of reading more, and more complex, writing but who still respond very positively to the compositional approach taken in this resource.

Another factor in the decision to compose the manual as I have is English fluency. A significant fraction of the students in these institutions are EAL learners, and while some of these are fluent in English, many are distinctly less than fluent, and many of these are not looking to use an already

[15] These are: Simon Fraser University and its partner institution Fraser International College, Douglas College (a large public college), Alexander College and Columbia College (private colleges).

challenging course as an opportunity to practice English. The challenges of higher education in a second (or third) language might seem self-evident, but the question has been studied, and studied specifically with CT (Floyd 2011). This was prompted partly by suggestions that non-Western students, in particular Chinese students, may be ill-suited to college-level CT education compared to Western students due to an education in their home countries that ignores or devalues or even suppresses the kind of questioning, inquiring, and flexible thinking seen as conducive to CT (Atkinson 1997). Although anecdotal evidence (conversations with students) suggests that some Asian educational systems do indeed place a greater emphasis on rote learning and mechanical skill mastery than many Western systems, if one is trying to explain why some non-Western EAL students struggle in an English-language CT course, their language disadvantage has to be a leading candidate explanation. Floyd reviews some general evidence that "most complex cognitive functions appear to be negatively affected by operating in a second language" (2011: 290) before reporting the results of her experiment, in which matched Chinese students took English- and Chinese-language versions of the *Watson-Glaser Critical Thinking Appraisal*.[16] Test scores and participant interviews both confirmed that, as hypothesized, thinking critically in a second language is more difficult than doing so in one's first language.

While this finding is hardly surprising, it's nonetheless an important one to keep in mind when designing a course to be taken by international students. It calls for caution with sentence structure and the overall volume of text, and also with idiomatic language and culturally specific references and examples.[17] Of course I need to balance the setting of this parameter against the goal of using examples and problems that are as realistic and meaningful as possible. These tend to be where the most challenging language is.

[16] This test is also widely used outside of academia, e.g. in business and law, in part because it is easy to administer and score. There are serious questions about the meaningfulness of scores on this assessment (Possin 2014).

[17] It's remarkable how easy it is to not notice how the idioms one uses would challenge an English learner, a point impressed on me once when I realized, in response to students' confusion about an exam question, that when an alarm *goes off*, it does the opposite of what one would have thought based simply on knowing the individual words.

Part 3. Course Design: Contents

Introduction

In Part 3, I discuss some of the most important things that students in this course learn about and learn to do. This is not a complete tour of Units 1-10, though I address most units in some way. I allocate attention based roughly on what content I think it is most important to explain and defend, given its role in the course, and that means that I mostly discuss argument, the course's most prominent tool for organizing and communicating reasoning. Since arguments are a standard part of academic philosophy, the treatment of argument analysis (recognition, reconstruction, and evaluation) is a major way in which this course represents philosophical practice to undergrad students who have taken few or no other philosophy courses. So this discussion addresses the fifth of the guiding questions laid out in the introduction, about how this course functions as a philosophy course. Then I investigate and respond to some actual and potential objections to a CT course that makes extensive, if not pervasive, use of argument. In addition to a sustained response to one provocative critic (Paul Thagard) of the traditional role of argument in CT, this section also has this project's longest excursion (Hugo Mercier and Dan Sperber) outside of the literature on the philosophy and pedagogy of CT. There is no shortage of that literature, but in section 3.1.4 I turn to a cognitive science perspective to CT and contemplate a possible novel objection.

The second main section in Part 3 concerns a loosely grouped collection of tools and topics that pertain to the Quantitative and Science designations of this course at Simon Fraser University, so this addresses the third and fourth of the guiding questions in the introduction, which were just how the course fulfills these obligations. I explain what aspects of deductive reasoning I think are important and useful to cover in a course that includes logic but is not a logic course – something I see as in the same spirit as the practically-minded logical instrumentalism of the *Port-Royal Logic* and the logic-for-consistency of Mill, though the specifics are quite different. Sections 3.2.2 and 3.2.3 discuss major elements of this course's contribution to scientific literacy in the broadest sense of that term. The closest thing to a consensus view among the mainstream definitions of *scientific literacy* since the mid-20th century is "a broad and functional understanding of science for general educational purposes and not preparation for specific scientific and technical careers" (DeBoer 2000: 594). "A functional understanding of science" includes acquaintance with the major domains, discoveries, theories, and development of modern science as well as a general ability to think like a scientist. This course contributes to the last element. Section 3.2 clarifies how these types of scientific thinking relate to the philosophical concept of abduction. Finally, while this course is not a mathematics course, it is more quantitatively demanding than many CT courses are and more than many students expect, despite its Quantitative designation. Bayes' rule is the centrepiece of this. I'll explain how successful students in this course genuinely *learn* Bayes' rule, not just as a little mathematical trick or opaque formalism, but as a fundamental and intuitive principle of rational thought.

3.1. Arguments

3.1.1. A Moderate Focus on Arguments

Argument is the single most important thinking tool I teach for organizing and communicating reasoning. Approximately 60% of the topics and sub-topics involve argument analysis or construction, although this accounts for only about 50% of class time. The topics that are concerned with arguments are:

Argument Topics	**Other Topics**
Unit 1: Concept and language of argument Unit 2: Deductive arguments and the concept of validity; basic categorical syllogisms; opposing valid arguments (a model of factual disagreement between good reasoners) Unit 3: Disjunctive syllogism; conditional arguments; using *modus tollens* to present an objection Unit 4: Inference to the best explanation and confirmation; disconfirmation with background assumptions (auxiliary hypotheses)	Unit 2: Universal generalizations and counterexamples Unit 3: Logical operators and conditionals Unit 4: Explanations and background assumptions Unit 5: Four errors under a broad concept of confirmation bias Unit 6: Concepts and basic rules of probability Unit 7: Bayes' rule and belief updating; tests and error types

Unit 8: Analogical arguments, particularly in support of normative conclusions Unit 9: Estimating a statistic (generalizing from a sample) presented in argument form Unit 10: Informal causal reasoning	Unit 9: Concepts of precision, accuracy, margin of error, and confidence level Unit 10: Causal explanations; controlled experiments; confounding factors; randomized controlled trials

So this is not simply a course on arguments, and students do not leave the course with the impression that CT just is the practice of analyzing and constructing arguments. Nonetheless the course is heavily invested in arguments as a tool for executing, organizing, and communicating reasoned persuasion. In this course, the arguments are at most paragraph-length compositions in which the inferential connections are established within a couple hundred words rather than an essay or book or speech. This is a somewhat unfortunate consequence of the practicalities of a course in which assessments must be based on exams and short assignments. Ideally students would be equipped to understand (and compose) larger, more loosely structured arguments, as well. But I see advantages to using short argumentative passages that require students to attend closely to the careful and precise use of inferential, logical, and other terms characteristic of particular types of arguments. This contrasts with the common student task of skimming longer texts to extract a main idea (see section 2.1.3).

The overarching message this conveys is that argument is the principal mode of rational persuasion. In Units 2-3, students get initial exposure to and practice with arguments as simple compositions in standard form. At this stage of the course, most students are hesitant with the concept of the underlying structure of argument, which may be instantiated by different examples. In these units, the concept of an argument pattern helps them understand validity, and the concept of validity helps them understand argument patterns (arguments are valid if they instantiate a valid pattern). Later, Units 4, 8, and 10 use extended argument patterns to analyze IBE, analogical, and causal reasoning. In the next section, I'll discuss the nature of the analytical exercises in these units.

3.1.2. Extended Arguments

One of the traits of a critical thinker I identify (section 1.2.2) is that they follow and construct chains of inference. While past experiments with exercises involving long chains of inference (with three or more intermediate conclusions) have convinced me that these are not the best use of our limited time in the term, I remain committed to establishing with students the basic principle of arguing by supporting a statement that in turn supports another. Assignment 1 (Appendix B) leads students through an exercise in constructing a brief deductive extended argument. Students encounter brief extended arguments to reconstruct in Units 4 (IBE), 8 (analogical), and 10 (causal). Examples here are written in a style that allows me to make argument patterns discernible for short reconstruction and evaluation exercises. This writing style is semi-realistic, natural and varied enough that it resembles the formal argumentative writing that may be found in serious journalism and opinion essays. But the deliberately structured passages also represent an idealized mode of writing with which students can learn, practice, and demonstrate different types of argument. Realistic examples would not have every statement in the full extended pattern explicit. I present maximally explicit argument patterns to explain the reasoning and then show students how typically some part of these patterns is implicit.

I represent IBE arguments as arguing from a judgement that some explanation is best (main argument) and optionally also in support of that judgement (sub-argument), analogical arguments as arguing from an analogy (main argument) and optionally also in support of that analogy (sub-argument), and method of difference style causal arguments as arguing from a judgement that some factor is the only relevant difference (main argument) and optionally also in support of that judgement. The merits of using these extended patterns differ somewhat between the three types but the general rationale is to accommodate their characteristic content when it is explicit and elucidate it when it is implicit without imposing implicit statements that aren't at least sometimes explicit in realistic examples.

Reconstructing analogical arguments as extended arguments is not standard. The pattern of an analogical argument used in CT education is typically (e.g. MacDonald and Vaughn 2019; Salmon 2013; Hughes and Lavery 2008) represented as a single inference:

- A and B both have features 1, 2, and 3.
- A has feature 4.

B has feature 4.

The philosophical literature has a wider variety of interpretations, partly reflecting the diversity of opinions but also reflecting the diversity of analogical reasoning, in science, philosophy, jurisprudence, everyday decisions, etc. Within philosophy of science, John Norton (2018) denies that there is any useful universal analogical argument schema on the grounds that all candidates are unacceptably vague or fallible. This view doesn't deny that analogical inferences constitute a real class of inferences, but only that their justification is to be found in adherence to a universal formal argument pattern; instead, analogical inferences are justified by rules that are either "local" (Norton 2018) or general (Bartha 2020). In Unit 8, I am forthright with students that the analogical reasoning as presented here is different from all the other reasoning tools in the course in being highly informal and intuitive. My goal is to give students a way to understand and discuss a common way that people speak and write – extending some feature from one item to another on the basis of their purported similarity – and this means working either with an argument pattern or a learnable analytic procedure that could well be an argument pattern.

Much of the philosophical literature has focused on what the right pattern is. Trudy Govier (1989) has oriented much of that discussion with a distinction between *a priori* and inductive analogies, with debate focusing in particular on whether *a priori* analogical inferences should be construed as deductive, which Govier denies but others maintain (e.g. Waller 2001). The conclusions of inductive analogical inferences are predictive and may be independently checked by empirical means, whereas *a priori* inferences are based on "an appeal to handle relevantly similar cases in relevantly similar ways" (1989: 142). The pattern above, lacking any additional implicit premise besides those ascribing properties to the compared items, appears to be generic enough that it could represent either inductive or *a priori* inferences. While many of my examples have a normative character and would fall under the *a priori* type of inference, many do not, and so I also need a pattern that is flexible enough to handle the full range of examples students are likely to encounter. The main problem with the standard generic simple pattern is that it is too reticent to directly acknowledge and accommodate the analogy itself, which is sometimes explicit and I think always important to establishing the style of argument. So I construe the analogical argument pattern as an extended argument with an analogy as the intermediate conclusion.

Main Argument

- A is similar to B.
- A has feature 4.

B has feature 4.

Sub-argument

- A and B both have features 1, 2, and 3.

A is similar to B.

The pattern acknowledges the analogy as a component of the argument, not just the style of the argument, something to be directly considered in an evaluation. With the simple pattern, the standard criticism of an analogical argument, that the analogy is weak or faulty, is awkwardly disconnected from the actual language of the argument. If analogies were always implicit in real-life analogical arguments, reconstructing analogical arguments without them might make more sense. But especially in cases where the analogy is surprising or contentious, it is often an explicit, prominent statement in the argument or its critique. For example, here's a critique of competing arguments over abortion: "abortion is not just like killing your toddler, just as restricting abortion is not just like enforcing bone marrow donations. Both analogies, put forth by abortion opponents and abortion-rights supporters respectively, fail by disregarding the natural and existential dependency of the unborn child upon its mother" (Bachiochi 2022). This critique directly attacks the "just like" claim, the analogy, at the heart of each argument.[18] In some cases the analogy is the *only* explicit statement offered in support of a conclusion. After some everyday examples like "this class is like watching grass grow", an example I use in class to help students reconstruct an argument around an analogy is taken from the concluding remarks of a 2007(!) scientific paper on the SARS-CoV family of viruses that calls the virus reservoirs in bat populations in close proximity to the wildlife markets of Southeast Asia a "time bomb".[19] We work through the various

[18] In some examples it is a little ambiguous whether the analogy is directly stated. Here is an example, with an inductive analogy, that I use in class: Baron and Klein argue that the insect brain supports subjective experience on the basis of claims about the neural functions that support consciousness in vertebrates and their contention that "structures in the insect brain perform analogous functions" (2016: 4900).

This example with an *a priori* analogy is more clear: "Jeanne Fromer, a professor of intellectual property law at New York University, said the [AI art-creation] companies may have a strong fair use argument. 'How do human artists learn to create art?' Professor Fromer said. 'They're often copying things and they're consuming lots of existing artwork and learning patterns and pieces of the style and then creating new artwork. And so at a certain level of abstraction, you could say machines are learning to make art the same way.'" The speaker asserts that the way machines learn to make art is similar to the way humans learn to make art. Excluding that assertion from the reconstruction – telling students that that statement is not actually part of the argument – would be a quixotic reading. And since the analogy is both supported by commonalities between the cases and used to support the "fair use" final conclusion, it must be an intermediate conclusion.

[19] Although the analogies in most of the examples I use are statements of similarity, I take metaphors like this one to be a type of analogy. This one may be treated as the statement that "the virus/market situation

ways in which the virus and wildlife market situation is like a time bomb (sub-argument) and then the authors' implied (main) argument that since we ought to prevent or prepare for the bomb's explosion, we ought to prevent or prepare for a virus "explosion" (outbreak) from a wildlife market. Reconstructing the argument around the explicit analogy, with the analogy at its heart, is a more natural way to analyze the passage than treating the analogy as a mere cue to reconstruct an argument without that claim in it. This approach actually aligns in this respect with Norton's "no formal schema" view in that his interpretation of analogical reasoning in science emphasizes what he calls the "fact of analogy", a conjectured "factual state of affairs that arises when two systems' properties are similar" that warrants an analogical inference (2018: 128).

It is unusual for someone to explicitly state every part of the extended analogical argument pattern. Usually something — typically either the analogy or the final conclusion — is implicit. So the best practice is to use the full pattern to help students grasp the content and mechanism of the argument, as well as its liabilities, and then use semi-realistic examples with implicit statements as a compromise between the artificiality of fully explicit standard form arguments and the unruliness of quoted real examples. (See examples in Unit 8 practice exams.)

In favour of the common non-extended pattern (without the analogy as a statement in it) is the fact that the evaluative question directed at the analogy is whether it is strong, not whether it is true, a question that seems applicable to an argumentative strategy but not to a premise in the argument. I teach that a strong analogy functions as the statement "A and B are very similar"; to call the analogy weak is to say that A and B are in fact not very similar, or that they are different in some key respect relevant to the conclusion. As I've said, I don't pretend that this argument pattern is more than a way of organizing common ways of speaking and of prompting evaluative reflection, since there is no general way to measure how similar two things are and no standard procedure to detect a relevant difference. The argument relies on a basically intuitive sense of when two things are similar enough to support the conclusion. But I believe this is an aspect of the informal nature of most analogical reasoning, or at least most of the analogical reasoning in the kinds of examples that concern me in this course. This method at least has the advantage of directly acknowledging the analogy at the heart of the reasoning so as to prompt the reasoner to confront the question of analogy strength. Once students have the claim that A is similar to B, it

---

is similar to a time bomb". Gentner *et al*. (2001) argue that novel (non-conventional) metaphors are processed as analogies in a process they call structure-mapping.

becomes easier for them to see that they must carefully weigh the significance of any dissimilarities they can think of between A and B.

I take a similar approach with IBE. The pattern involves an inference from the claim that some explanation is the best one to the claim that it's the true one. Presenting the pattern with these distinct statements, one supporting the other, helps students see how IBE works. Many students, unless directly presented with the two statements in a way that highlights the difference, fail to even distinguish between the normative judgement that a hypothesis is the best and the assertion that it is true. This conflation makes it impossible to properly understand IBE. The process of applying criteria of explanatory quality (e.g. simplicity) and considering alternative hypotheses supports a judgement that one explanation is best, which is why a "best explanation" claim is explicit in many arguments, particularly in scientific writing. The inference to truth is a further step. And we may be surprised — the best explanation isn't always the true one. Understanding this is essential to understanding the limitations of IBE as a non-deductive inference. The fully explicit pattern in which both statements appear captures all of this.

And yet, as with analogical arguments, it would be unusual for someone to explicitly state both the "best explanation" premise and the "true explanation" conclusion. Typically someone either states that some explanation is the best one and leaves it implicit that they tentatively judge it to be true, or, having given an *explanandum*, they conclude that the explanation is true by implicitly judging it the best. The former abbreviation is especially likely in cases where the "best explanation" premise is not merely asserted but supported in what I represent as a sub-argument that considers competing explanations. So again my approach is to use the complete extended pattern to show how this type of reasoning works, and to use semi-realistic examples with an implicit statement to develop students' awareness that these reasoning patterns are really present, and can really be used without being pedantic, in real-life formal writing. (See examples in Unit 4 practice exams.)

Some indication of the value of analyzing IBE with this extended pattern is this passage on IBE from the popular textbook *The Power of Critical Thinking*:

> Very often we may propose a theory as an explanation for a phenomenon, or we may have a theory presented to us for consideration. In either case, we will likely be dealing with an argument in the form of an inference to the best explanation. The conclusion of the argument will always say, in effect, *this* theory is the best explanation of the facts (MacDonald and Vaughn 2019: 358).

In the authors' presentation of the IBE pattern (44 pages before their discussion of it), the "best explanation" claim is a *premise* and the conclusion is "it is probable that E is true" (2019: 314) — the main argument in my extended IBE pattern. MacDonald and Vaughn don't notice that they give two completely different versions of IBE. What they are actually doing, without realizing it, is describing first the IBE main argument and later the IBE sub-argument. The "best explanation" claim *is* a premise, but it's *also* a conclusion: it's an intermediate conclusion in an extended pattern. IBE is naturally an extended argument pattern. They just don't acknowledge this by laying out or using the full pattern, as I do.

Reading and reconstructing, or writing, argumentative text with this attention to the underlying structure is challenging for the many students, both EAL and native English-speaking, who have rarely if ever been asked to process or produce this sort of highly deliberate writing. Along with the argument construction tasks of Units 2-4 and the related evaluation problems in Unit 5, the argument analysis tasks of Units 4, 8, and 10 represent a significant commitment of the limited resources of time and student attention. So this raises an important question: Is this a good use of these limited resources? As I briefly reviewed in Part I, there is a long tradition of informal logic that predates and then interacts with the CT movement, and the major focus of this tradition has been arguments and fallacies. Contemporary CT has inherited this focus, a view summed up by Anthony Blair, who declares that "the centrality of arguments to the art of critical thinking is unquestionable" (Blair 2021: 95). The attention I give to arguments, and the argumentative competence I incorporate into my definition of CT, is within this tradition, although even more than Blair, I stress that "critical thinking has other objectives in addition to appreciating arguments". But if there are reasons to question devoting approximately half of a CT course to argument-related topics, I should address them. In the next two sections, I consider some actual and potential criticisms of a CT pedagogy that gives this prominent place to argument.

3.1.3. Thagard's Challenge to the Utility of Arguments

Paul Thagard (2011) advances a contrarian view of teaching CT. His primary target is the psychological theory, which he believes to be the basis for most philosophers′ approach to teaching CT, that real human inferences are underpinned by arguments, that rational beliefs and decisions are produced by good arguments, and that defective beliefs and decisions are produced by fallacious arguments. This theory is false, he claims, because inferences are completely different from philosophers′ arguments: ʺHuman inference is actually parallel rather than serial, multimodal rather than just language-based, and as much emotional as cognitiveʺ (2011: 156).

Studying the traditional fallacies is mostly useless and should be replaced by psychologically informed attention to human error tendencies, foremost among them motivated inference, work "more akin to psychotherapy than informal logic" (2011: 157). Thagard's criticisms have some merit, and some features of this course reflect similar views, but I believe that his case against a CT focus on arguments goes too far and that some of his prescriptions for CT pedagogy are misguided.

Thagard alleges that argument-focused CT falsely "assumes that inferences are based on arguments" (2011: 154), but this allegation does not accurately characterize my view. I make no empirical assumption that people reach their beliefs and decisions by constructing private arguments which they sometimes communicate to others. I treat arguments as a tool for organizing, inspecting, and communicating reasons for belief to others, a process that forces the arguer, and prompts the reader/hearer, to be explicit about those reasons and the nature of the support they give. Our inferences *can* be based on arguments, others' or our own, when we take our mind through the process of following or constructing an argument. Thagard offers nothing to challenge the notion that this is possible or beneficial. The fact, granting that it is one, that humans do not instinctively draw inferences in this manner doesn't establish that they cannot do so. Humans don't instinctively use written language, either. Learning to read and write is difficult and in a certain sense unnatural. But we can learn to do so and benefit immensely, on our own and in communications with others, by doing so. I devote significant attention to arguments because inferences based on arguments can be better or worse, depending in part on the quality of the guiding argument. This is the main justification for studying them.

Thagard presses his case with the claim that although the limitations of working memory constrain conscious thoughts to be serial, "the formation of these thoughts is the result of a massively parallel process that integrates many sources of information" (2011: 155), a process alien to the serial, linguistic nature of argument. This is a puzzling complaint. Thagard may be overly concerned by the convention of labeling premises "Premise 1", "Premise 2", etc. I do this occasionally for exam questions when I need to be able to clearly refer to specific premises in instructions. Generally I use simple bullets so as not to create the impression that the premise order matters. The argument is an abstract structure that is built out of statements with connections to each other. It is not a sequence of operations. For example, an intermediate conclusion may be seen as carrying an argument "forward" from a sub-argument, or as referring "back" to the support for a premise in the main argument. And as Thagard notes, an argument may guide belief "forward" from acceptance of premises to acceptance of conclusion or “backward” from rejection of a conclusion to the obligation to reject at least one premise. At most this structure is linear, not serial, though when multiple sub-arguments are present, it’s not even linear. It’s just a linguistic structure that organizes interrelated statements.

Thagard is impressed by the fact that the brain “carries out many processes simultaneously” and “involves around 100 billion neurons firing asynchronously” (2011: 154-155). But this cannot show anything useful about the utility or nature of arguments. No one thinks arguments are “in the brain” — they are built and cognized by brains, subject to the limitations imposed by sensory, motor, and linguistic processing, attention and working memory, etc. The relevant question is whether it’s epistemically and pragmatically useful for brains to make use of these thinking tools. I think it sometimes is, along with other abstract structures whose utility for CT Thagard’s argument would undermine if successful. For example, Bayes’ rule seems to be sequential in whatever weak sense that arguments might be said to be: the rule guides a reasoner from a few pieces of information (prior probability, likelihood, and marginal probability) to a posterior probability. In fact, mainstream neuroscience takes quite seriously the possibility that brain mechanisms in some way instantiate or operate according to Bayes’ rule (Doya *et al*. 2006). Either way, this has little to say about the utility of teaching students Bayes’ rule as a tool of thought. A non-Bayesian brain wouldn’t undermine the advantages of comprehending Bayes’ rule; a Bayesian brain wouldn’t establish them.

Thagard is right to be dissatisfied with a pedagogical strategy that hopes to improve people’s thinking by teaching them fallacies, although there is more to be said than his dismissive claim that many traditional fallacies are “arcane and rarely committed by people in real situations” (2011: 154). There surely are many arcane, perhaps even mythical, fallacies among the lists that have been compiled over the millennia, as well as implausible applications of fallacy concepts. For example, is it truly helpful to tell students that an advertisement featuring a celebrity athlete happily drinking cola brand X is an argument for the conclusion “You should drink cola brand X”, and that it commits the fallacy of arguing *ad verecundiam* (in the Latin, of course) (Woods *et al*. 2000: 33)? I think this is inviting students to dismiss CT as bizarre and irrelevant.

Teaching CT by working through a catalog of fallacious arguments to be avoided gives an undue primacy to fallacies, although I wouldn’t go as far as even Hitchcock does in his case against fallacies: “it makes more sense to teach the analytical apparatus for correct reasoning [e.g. legitimate appeals to authority, generalizations from well selected samples] (and to let the fallacy fall out as a kind of deviation) than to begin with the fallacies” (2017: 406). My usual approach is to develop an account of the good reasoning and then to show some errors that people commonly make in using it. Units 4 and 5 (IBE/confirmation and associated biases) are the largest example of this. Smaller examples are some formal fallacies (Unit 3), the gambler's and conjunction fallacies (Unit 6), and the fallacies, if they can be called that, of a weak analogy (Unit 8) and of ignoring another relevant difference (Unit 10).

However in some cases it really can help to explain the good reasoning with the aid of carefully chosen examples of a fallacy, though the *fallacy* terminology may not matter. Examples illustrating the fallacies of ignoring a base rate, a biased sample, or ignoring a confounding factor, are effective ways of helping students to understand Bayes' rule, good sample selection, and good causal inferences. When an example is presented in the right way, students intuitively grasp that there is an error but usually lack the ability to say precisely what it is. Working out what the error is and how to describe it using the new concepts they are learning is an effective way for students to learn what the good reasoning is. For example, a good sample is selected in a way that is *not biased* (nothing about the selection criteria or procedure is correlated with the feature being measured). Sample bias is the primary concept here so it can work to teach the good thinking by teaching the bad. The question in any particular case in which one must decide whether to teach CT in this way is whether the fallacy actually represents a significant human error tendency that students are likely to encounter or commit. If not, this is likely a poor or at least limited strategy. Thagard's hostility to the traditional emphasis on teaching fallacious arguments goes too far. There is a place for it, and while in many cases it may be possible to present the good and the fallacious reasoning without the apparatus of an argument, Thagard's complaints do not show that it is important or better to do so.

Thagard places motivated inference at the top of his list of real human reasoning errors that a CT course should ensure that students understand. I lightly incorporate the *motivation* aspect of the error into classroom discussions of biased search for evidence, *ad hoc* explanations (protecting a hypothesis by giving an otherwise unsupported excuse for disconfirming evidence), and imprecise predictions (which can function to make a test of a hypothesis too easy). But Thagard's approach risks setting students loose to indulge in psychologizing, responding to faulty or absent reasoning with eager "identification of conscious and unconscious goals that can explain why people are inclined to adopt beliefs" (2011: 157) rather than criticism urging better or more thorough use of available evidence, perhaps organized in an argument as a way to help assess its objective significance. Warning that everyone is biased in ways they "are rarely consciously aware of" (2011: 157) licences or pushes students to go beyond a useful awareness of the real human tendency to make improper use of evidence to a blanket suspicion of evidence-based persuasion that is hardly better than the paranoia of the fallacy hunter.

Thagard's proposed tool for having students do this is an innovation he calls a "cognitive-affective map", an impressionistic diagram tracking valanced connections among the evidence, beliefs, and overt and suspected motivations that lead someone to hold the position they do. This is to aid students in "analyzing the emotional character of conflicts", which "arguments conceal" (2011: 161-162). I think there are good reasons for people to try to understand one another's feelings and motivations, whether they are in or merely seeking to understand a conflict. This message is compatible with, even complementary to, the goals of CT. But I do not sign on to Thagard's view of CT as psychotherapy, and I do not agree that argument analysis is fruitless without these speculations on the emotions and unconscious motives behind the presented evidence. The advantage of an argument is that it lays out the evidence that the arguer asks others to accept as supporting the conclusion. It allows all concerned to see what factual claims are being called upon, where people's beliefs align and diverge, and how people agree or disagree on the degree of support provided by the evidence. Of course arguments are not the only tool that does this. A more or less explicit use of Bayes' rule that prompts people to isolate and acknowledge different estimates of the relevant probabilities also does this. If people interpret the same evidence very differently, it may be a path of dispute resolution to widen the discussion to investigate unstated values and commitments. But the foundation of that kind of progress will be a clear understanding of where and how people agree and disagree on the evidence and what it means. And that means arguments and argument analysis remain vital tools not only for arriving at rational beliefs but for having productive disagreements.

Only in its conclusion does Thagard's case against arguments grudgingly acknowledge the role of arguments in "communication of evidential considerations that go into good inferences", a role that is "psychologically indirect but socially significant" (2011: 164). So there is some role for arguments in CT, after all. It's a striking concession after earlier taking aim at the idea that people's beliefs are sometimes based on arguments. He appears not to contest the idea that people sometimes do form beliefs based on communication with others; and he grants an important role for arguments in communication. It seems he must grant that people's beliefs are sometimes based on arguments. If Thagard's original contention is merely that people do not always base their inferences on arguments, this is banal and undeserving of the radical character of his paper. But the stronger and more interesting claim, that a CT course with a robust treatment of argument analysis, informal logic, and a few well chosen fallacies is misguided and psychologically naive is not supported by Thagard's own claims, as we can see by reconstructing his argument. However often people do in fact formulate their beliefs by inspecting the organized presentations of evidence called *arguments*, they *do well* to do so. And most people would do well to do so more often and more competently than they do. So devoting approximately half of a CT course to arguments and associated errors is a reasonable design decision.

3.1.4. Mercier and Sperber's Evolutionary Theory of Argument

Mercier and Sperber's (2011) "argumentative theory of reason" claims that the main biological function of human reason is to produce arguments that convince other people, though evaluation of arguments is also part of its function. The theory is a contribution to the field of evolutionary psychology, not to the mainstream philosophical discussion of CT pedagogy. I'll discuss it here because it's an important recent theory concerning the human phenomenon of argument, some parts of which might reasonably be interpreted as having significant implications for how or even whether arguments should be taught in a CT course. Thagard rejects M&S's theory because he sees argument as a "cultural development" (2011: 164) that actually serves *more* helpful purposes than the biological theory seems to allow. If even an argument skeptic like Thagard thinks that M&S give too little credit to arguments, it seems the argumentative theory will be a dismaying one for anyone hoping to improve students' thinking through argumentative competence. In fact, the implications are mixed. Like Thagard, M&S claim that inferences are not based on arguments; unlike Thagard, they claim that arguments are directly connected to our inferential processes, that arguments are the "outputs" of our "intuitive inferences" (2011: 58). This raises the disconcerting prospect that an argument is a kind of confabulation, subjectively understood as a product or expression of deliberation but in fact a product of sub-personal arational mechanisms. The argumentative theory implies that argument construction by an individual is an unpromising route to truth. However it also suggests an evolutionarily rooted rationale for teaching argument evaluation and promoting the communicative role of arguments.

In outline the M&S argumentative theory is this. Humans evolved to make use of and rely on extensive information exchange with one another. When a communicator wants to impart information that an audience is likely to reject because it conflicts with a prior belief, they must either convince the audience of their general trustworthiness or try to leverage whatever facts they already agree upon, in other words, argue. M&S review a range of experimental evidence showing that people are much better at argument production and evaluation in dialogic contexts and in groups than they are in abstract, socially disconnected exercises. This supports the central claim of the theory, that argument production is part of a social function of reason. The theory predicts pervasive confirmation bias, or my-side bias, in the production of arguments, though not in their evaluation, although if evaluation goes beyond critical scrutiny to become a full counter-argument, the theory predicts my-side bias in its production. Biologically, this is a feature, not a defect, of reason. M&S review evidence that people reason better for hypothesis testing and categorical syllogisms when they operate in an epistemologically defensive or critical mode rather than alone or with others who already closely agree with them. In contexts where people "disagree but have a common interest in truth — the confirmation bias contributes to an efficient form of *division of cognitive labor*" (2011: 64) where some parties put forth ideas and others test them, and collectively the group finds its way toward truth (or at least, toward whatever sorts of truth it is evolutionarily beneficial to find). A variety of forms of motivated reasoning function not as the self-deception they appear to be but "proactively", to be ready when challenged by others.

M&S allow that a tiny minority of "freakish" people internalize the division of cognitive labour and subject their own inferences to genuine scrutiny, but these are the exceptions that prove the rule that people instinctively apply their reasoning skills in ways that, individually, are not truth directed.[20] They claim that the theory does "not paint a wholly disheartening picture" from an epistemic perspective (2011: 72-73). Is the theory, if broadly correct, un-disheartening enough that a CT course with a significant focus on arguments pursues a productive educational strategy, or is this naively optimistic? If arguments are just *post hoc* justifications and a mechanism of social manipulation, and more generally if arguments are *outputs* of our inferential processes rather than a tool of inference, this suggests that teaching arguments may be unhelpful or even counterproductive. The best approach may be to warn students of their true deceptive nature.

The treatment of arguments in this course addresses some of the concerns that arise from M&S's theory of reasoning. The argumentative theory suggests dim prospects, for most people, for the private pursuit of truth through arguments. The course mostly does not present arguments as this sort of tool. Units 2 and 3 on deductive reasoning are a partial exception. In these units, arguments are a tool for noticing implications of beliefs and requirements for certainty. Students at this point in the course are just developing the general idea that statements can be related to one another such that the truth of some statements entails the truth of another statement, or that the falsity of one statement entails the falsity of at least one other statement. Developing the habit and skills

[20] M&S's freaks are reminiscent of Wellingham's (2010) rare minds capable of internalizing and appropriately deploying general CT skills. See note 4.

of maintaining logical consistency in one's beliefs is a truth-directed project insofar as believing a contradiction is a guarantee of believing a falsity, and failing to believe something implied by already justified beliefs is a missed epistemic opportunity. While limited, the epistemic rationale of this private use of argument is not undermined by the argumentative theory. Confirmation bias may still corrupt deductive reasoning by the unwarranted acceptance of false premises, but training in the construction and completion of these arguments puts tools in place for students to notice their rational commitments and perhaps question or weaken the beliefs that have these entailments.

It's also in these units, and the deductive component of Unit 4, that students try their hand at communicative argument. This is Assignment 1, a composition framed as a simulated academic, professional, or personal communication (see Appendix B). The argumentative theory predicts that argument composition will be easier in contexts that activate people's reasoning abilities in order to convince others. In Part 1, I noted Trudy Govier's (1999) concept of a "minimal adversariality" that is intrinsic to argument: to argue for a conclusion A is to argue against NOT-A, a position that might be held by others who would thereby be one's opponents (Govier reconsiders some aspects of her model in (2021).) The evidence M&S draw on suggests that the minimal adversariality of an argumentative composition is not by itself enough to engage people's natural arguing skill, so Assignment 1 puts students into a disputational frame of mind by prompting them to argue against some hypothesis that they can see some others do or might believe (a judgement they also discuss with me in a brief interview in courses where enrollment and time permits). Since they identify a hypothesis to argue against, they begin by conceiving of a potential opponent's position. They must also acknowledge a background assumption upon which their prediction is based and confront the possibility that it's false, as an opponent who defends the hypothesis might allege. Consistent with the M&S theory, this assignment format engages students far better than a less specific instruction to construct an argument for any conclusion, with tends to result in more forced and stilted compositions.

Regarding the problem of my-side bias, the M&S theory counsels us not to hope for more than a "painstakingly acquired ability to exert some limited control over one's biases" in some elite reasoners (2011: 73). On one hand, this discouraging upshot of the theory doesn't warrant total defeatism. The argumentative theory doesn't entail the outright futility of helping students learn to notice and compensate for, perhaps even sometimes suppress, my-side bias. Maybe this will be painstaking, and maybe it will be rare, but if M&S are right that it is an acquired ability, this underscores the value of a CT course that gives students some basic concepts with which they can understand and improve their own thinking, even if is only the rare hero who succeeds in overriding their instincts. One might see similarly dim prospects for people's solid competence in math or chess or piano or tennis, but it would be a mistake to simply write off these endeavours (though there may be other reasons not to subject children to intense training in these things unless they're strongly attracted to them). Skilled performance is sometimes possible and valuable, and different types of instruction can be more or less effective. Moreover all sorts of behaviours rooted in evolved psychological mechanisms are resistant to change but modifiable with the right tools, such as treating binge-eating with cognitive-behavioural therapy (Lindardon 2017). On the other hand, the argumentative theory clearly has implications for a CT course if one takes it seriously, as I do. I've said above that I don't see psychotherapy as the right tool for this problem. Though M&S call it an ability, exerting control over one's biases would seem to have a significant dispositional component, and I've already said (section 2.1.2) that this course has limited ambitions to reshape students as people. But teaching students how confirmation bias is a real epistemic hazard involves promoting only very generic values and character traits such as a concern for evidence and truth. The more specific disposition to be unbiased in one's reasoning is difficult to disentangle from the ability. The strategy of trying to teach students to exert control over biases as they arise by teaching students what common biases are and how to recognize them is roughly what Tim Kenyon and Guillaume Beaulac (2014) call the "intuitive approach to teaching debiasing". In line with M&S, they hold out little hope for this approach, which they see as doomed by the phenomenon of "bias blind spot", in which someone may accept that people in general exhibit bias, and even that they themselves do on other occasions, but still convince themselves that they in *this* judgement on *this* occasion are unbiased. Kenyon and Beaulac hold out more hope for "collective debiasing strategies and infrastructure" (2014: 343) and in some cases for individuals to learn to incorporate "nudges" into their own lives to systematically manage rather than directly self-suppress bias.

The present design of the course allocates modest space to the topic of biases. The topic is tied most directly to IBE and confirmation reasoning. The approach I take, illustrating some forms of confirmation bias as failures or misapplications of this reasoning, and teaching students to diagnose these, could be accused of responding to the glimmer of hope in M&S's account with the hopeless "intuitive approach". However the principal aspiration of this unit is not to transform students into M&S's unbiased freaks. I do hope to help some students become more freakish, but I think a realistic ambition, still valuable, is to equip students with some concepts to better understand the reasoning of the preceding and following units, and to articulate constructive criticism of some forms for confirmation bias when they encounter them. Most of the forms of confirmation bias that I cover do not render the entire piece of reasoning a worthless deception that could not be improved. Evidence collected in a biased manner may be incomplete but still useful when supplemented with additional evidence. *Ad hoc* explanations of disconfirming evidence may be overeager but at least illustrate the importance of considering background assumptions – good judgement consists in discerning the difference between a background assumption that is a significant liability to the inference and one whose truth is secure enough that rejecting it in response to disconfirming evidence smacks of desperation. Imprecise predictions are insufficient to function as a meaningful test of a hypothesis but they can be improved with consideration of what prediction would have a high enough chance of truth, were the hypothesis true, that its falsity is informative, but a low enough chance of truth, were the hypothesis false, that its truth is informative.

M&S predict, and point to evidence, that people are less biased in evaluating arguments than they are in producing them. This gives an evolutionary underpinning to something like John Stuart Mill's (1859) classic opposition to censorship and defense of free speech, also interpreted as a defense of the debate process (Woods *et al*. 2000). The consolation M&S offer is that in group reasoning situations "*truth wins*" (2011: 72) through the process of individual arguers trying to justify their prior conclusions and others exercising "epistemic vigilance" (2011: 60) by subjecting those arguments to critical scrutiny, including the detection of bias. Again this is in line with Kenyon and Beaulac's aim of debiasing at the "collective" level. While a single course does present limited opportunities to explore collective (e.g. institutional) strategies, equipping students with the concepts and skills to recognize and articulate specific examples of specific types of bias when they see them, even in brief narratives, is a foundation for this. It will certainly be helpful, if not necessary, to understand what tendencies we're trying to suppress. Students aren't learning to do this for the sake of privately dismissing others' beliefs, but so they can participate in discussions about poor and better uses of evidence in support of beliefs, ideally with others equipped with the same concepts. Of course using evidence in support of beliefs is arguing. Understanding the forms of confirmation bias goes hand in hand with understanding the forms of confirmation argument that are stronger when they're not infected by these biases.

In devoting significant attention to the reconstruction and evaluation of arguments, this course aims to equip students who are exercising epistemic vigilance with concepts that help them understand what to expect and what to demand from others who would convince them to hold new beliefs (or maintain existing ones), and that help to establish mutual comprehension and productive adversariality. It hardly needs to be said that, for a variety of reasons, often truth does not win even in group reasoning situations, as in individual ones. But the purpose of this section has been to address potential objections to teaching arguments that arise from the argumentative theory. The underlying view of this course is not that argument analysis skills are all people need to flourish as critical thinkers but that they *are* valuable CT tools, along with other skills and dispositions.

3.2. Deduction, Science, and Probability

3.2.1. Deductive Reasoning

The logical ambitions of this CT course are modest. This course is normally offered alongside a separate introduction to logic course, and I aim to duplicate very little of that content. Units 2-3, the first units to which I devote significant class time, are on logic topics, and Unit 4 has a connected section on disconfirmation. In this section, I'll outline the role of logic in the course.

Peter Geach distinguishes between two senses of *logical thought*: 1) "thinking that is being commended as orderly, consistent, and consequent, whatever its subject-matter", and 2) "the thinking of logicians about logic" (1979: 5). Correspondingly there are two senses of teaching logic: 1) helping people to make their thinking more orderly, consistent, and consequent, and 2) training people to be logicians. (By *thinking consequently*, Geach means that someone notices and believes the implications of their beliefs, and if they find they don't believe one of those implications, they accept that at least one of the beliefs that implied it is false.) I do not aim to accomplish the second goal. I do aim to make some highly practical advances of the first kind. I try to help students acquire simple logical skills for clarity in everyday thought and communication, and I use certain features of the logical exercises in these units to set up the projects in the subsequent units.

Most students arrive in this course knowing *logical* to be a loose synonym for *reasonable* or *sensible* and that's it. The first objective in going beyond that very general understanding is for students to appreciate that there is a special way of being reasonable that has a character akin to that of mathematical operations, a black-or-white character reflected in some key binary oppositions: true/false, in/out (of a category), consistent/inconsistent, valid/invalid. Some forms of black-or-white cognition evidently come quite naturally to students, including the first I discuss in class in Unit 2, the notion of a category that either includes or excludes every thing in the world.

Even given an illustrated option of endorsing a kind of family resemblances notion of *Living Things*, a large majority of students favour a category with a sharp border that forces an in-or-out judgement on a controversial case like a virus (see Appendix D).

And yet the formal nature of exercises in this section is an immediate challenge for students. As the psychologist Philip Johnson-Laird remarks, “human reasoning is not simple, neat, and impeccable” in part because it so strongly “tends to exploit what we know” (2010: 18249). Drawing students’ attention to the underlying pattern of a statement, to the very idea of a statement’s underlying pattern, is one of the key tasks in these units. It takes effort and practice to see past the subject matter, which automatically activates background knowledge, to an underlying pattern that is common to other examples and that can be captured in a generic sentence or diagram. Helping students notice what is common in statements and arguments with different subject matter is essential for helping them do even the limited logical tasks of these units, the construction of a few types valid arguments and counterexamples, and a basic demonstration of invalidity with a counterexample (a scenario in which the premises are true and the conclusion is false), a skill Johnson-Laird calls “the heart of human rationality” (2010: 18249). One reason I do these units at the beginning of the course is that this aspect of the topic provides relatively simple, neat examples of identifying and working with “deep structure”. That a CT course should be designed to make salient the deep structure of diverse examples was the second lesson I drew from the CT generality debates (section 1.3.3). I see the logic units as serving an important function in establishing this as a general way of thinking that will apply in other cases where the deep structure is something else, such as an inductive argument schema, a conditional probability, etc.

For anyone with a personal or professional commitment to the traditional philosophical conception of logic, the psychological literature on the topic contains some startling assertions. “In standard logic, MP is considered a valid inference, whereas AC is fallacious”, write Niki Verschueren *et al.* (2005: 108), discussing the role of working memory in everyday reasoning, but “this valid/invalid distinction does not apply to everyday reasoning”. They explain how the distinction does not apply using a case described by the “mental models” theory of reasoning, associated with Johnson-Laird. *If someone pulls the cat’s tail, the cat gets angry*. If people process this case with a single, restricted mental model in which the tail-pulling is the *cause* of the cat’s anger, they endorse both the MP and AC inferences (agreeing with the conclusion when the premise is stipulated). But for many people, “active consideration of the problem content will lead to an automatic activation of additional relevant information from long-term memory” (2005: 108) and they endorse *neither* inference. MP becomes a bad inference because people think of "disabling conditions" (e.g. sedated cat); AC is now a bad inference because people think of "alternative causes" (e.g. arrival of a dog). Untutored reason exploits knowledge in two very different ways in this example. The disabling condition calls into question the truth of the conditional — it proposes that tail-pulling is, in fact, not sufficient for cat anger. The alternative cause describes a counterexample scenario in which the cat is angry though no one pulled its tale.

I understand Verschueren *et al*.'s claim to be not or not just (plausibly) that most everyday reasoning is non-deductive, but that "standard logic" is in some way psychologically naive. While I won't take up that claim here, I do agree that this type of example shows that it's unnatural and difficult, at least at first, for people to separate factual and support questions even when asked about support only. But that is an essential CT skill because it allows people to have useful discussions about what would or could support some contentious claim before turning to see whether mutually agreed upon facts do this. This is bigger than a logical skill but deductive exercises such as those in Units 2-3 are well suited to developing it because the question of whether premises fully support the conclusion can be answered without reference to background information. Of course there are only two ways to order two topics (deductive and inductive reasoning), but some prominent CT textbooks place inductive before deductive reasoning on the grounds that "students are more familiar with inductive arguments" (Salmon 2013: viii). Familiarity and relatability are a reason to have a general introduction to arguments that uses everyday examples of both deductive and non-deductive arguments, as I do in Unit 1. And they're a reason to focus on just a few basic deductive argument patterns and to connect the early deductive reasoning with other everyday thinking skills such as counterexamples (Unit 2). But with those precautions in place, they're not a reason not to take advantage of using deductive reasoning to crystalize the fact/support distinction.

As I said in section 1.2.3, I think there are fewer unambiguous examples of logical fallacies – inferences where there is no reasonable suspicion that an implicit assumption (e.g. an implicit biconditional) is doing important work or that the inference is intended as confirmatory rather than valid – than a deduction-heavy CT course would suggest,[21] and I don't drill students with exercises

---

[21] This is roughly the same judgement that Finnocciaro (1981) reaches, although his judgement is based mostly on a survey of the meager selection of real examples in logic textbooks. Govier (1987) suggests that textbook authors had simply been lazy in collecting real examples. While I don't deny that people ever construct or are deceived by genuinely invalid arguments, I think a CT course makes itself more relevant to the navigation of everyday discourse by helping students see that such cases are invalid as written when interpreted deductively, though their true nature may be indeterminate, at least without further probing.

focused on the detection of invalidity. The main goal of the logic units is to help students make their thoughts and communications more orderly, consistent, and consequent. To that end, the following are major objectives:

1) Students gain some appreciation for the careful and precise use of language, including many common words and phrases they use all the time without much thought. This includes attention to a few things that might not normally be regarded as essential but that I believe have great value for building communication skills. For example, students learn to understand and use conditionals in the …*only if*… form in addition to the standard *If*… *then*… form (which helps in explaining the difference between necessary and sufficient conditions). And I think that *only* is such a common and useful word that I teach it alongside *All* and *No*, and students learn logical equivalences with all three.[22]

2) Students gain skills for analyzing a reasonable diversity of linguistically complex generalizations (not just "All dogs are mammals", etc.) and for representing them with simple Euler diagrams that give concrete expression to the relationship between categories.

A (video) example of this I use is from the animated comedy *South Park*. Randy has stayed up all night cooking a gourmet breakfast for his family. His wife Sharon comes downstairs with the kids and, seeing this, concludes that he has been watching cooking porn, since "every time he watches those shows, he stays up all night cooking". This is formally invalid as stated. But is it really best construed as a logical error, or does Sharon simply mean but not say that every time, *and only times*, that Randy watches cooking porn shows, he stays up all night cooking? I use the example because the video is funny and engaging and shows that the pattern wasn't just dreamed up by philosophers, but I've found that denouncing it as bad reasoning, as an example of something to be detected and banished on the grounds that it's not truth-preserving, is a less-than-compelling demonstration to students of the value of logic. I think the main value of the exercise is for students to recognize that, in order for this argument to be valid, it would need a separate *Only*… claim (or a different *All*… claim), which may be implicit here, and also may be doubted. And it's important to proceed efficiently to probability, where students learn that the really useful questions to ask in this case would be: what are $\boldsymbol{P}$(WatchShows), $\boldsymbol{P}$(Cook|WatchShows), and $\boldsymbol{P}$(Cook|¬WatchShows)? If Randy cooked and $\boldsymbol{P}$(Cook|WatchShows) = 1 (Sharon's premise), this *is* relevant to whether Randy has been watching cooking porn. But she wouldn't have deductive proof unless $\boldsymbol{P}$(Cook|¬WatchShows) = 0.

[22] Geach relates an episode in which a student answered an exam question that used a statement beginning with *Only*. They were "apparently of the firm conviction that any logically well-dressed proposition will wear on its head either an 'all' or a 'some' or a 'no'; for the answer began with the unforgettable sentence 'Since we are doing Logic, we must assume that "only" means "all"'" (1979: 12). Many students will indeed indicate that *only* means *all* when the question is first put to them as a problem of logical equivalence. Geach's student appears to have reached the exam without any useful practice with *only*. That does not happen in this course.

3) Students get some initial practice with arguments after the concept is briefly introduced in Unit 1. They distinguish premises and conclusions and begin to see that arguments can be good or bad according to particular standards, depending on their form.

4) Students especially get a chance to practice constructing *modus tollens* arguments in exercises that combine structured thinking with consideration of acceptable conditional premises.

5) Students practice articulating the relevant properties of counterexamples and distinguishing things that are (or would be) counterexamples from things that are not. They also work through alleged counterexamples as part of a disagreement structure of opposing valid arguments, a dispute between two good reasoners in which one person infers from a universal generalization that X lacks some property, and the other person infers from their belief that X has this property that the universal generalization is false (that X is a counterexample) (see Unit 2.iv of the course manual for the *terrorism* case, or Appendix D for the *meat* case).

6) We lay the foundations for the probability units by establishing the meanings of negation, disjunction, and conjunction. Discussion in class includes examples where the connectives have important and easily overlooked significance. For example, disjunctive headlines or key findings can seem more sensational than they are — some can be nearly uninterpretable — by linking an uninteresting, probable fact to an interesting, improbable one. We see how important these words can be in legal, policy, and contractual texts. A vivid example comes from a draft of Simon Fraser University's policy on sexual violence and misconduct which defined *sexual misconduct* to include "distribution of sexually explicit photographs or videos of a person without their consent and with intent to cause distress". It's an eye-opening moment when students reflect on this and realize that it invites the unwelcome defense that an instance of such distribution, though lacking consent, was not intended to cause distress. (The final version of the policy deleted the conjunct after I submitted a query during the feedback period.) It's even more eye-opening when I show a recent example of a law in the UK formulated in this way that led to exactly this painful outcome. Even then, it takes some work for many students to see that a law against **P** and **Q** (e.g. image distribution and lack of consent) prohibits an instance of **P** and **Q** and **R** (image distribution and lack of consent and intent to cause distress), but that including **R** in the law substantially increases the burden of what a prosecutor (or university committee) must prove, perhaps for no good reason of justice.

This is not Richard Whately's CT course. Deductive reasoning plays an important role but it's not presented as the model of all good reasoning. There is plenty left for graduates of this course to learn in an introduction to logic. But this limited set of deductive and related skills deserves its place early in the course both for its value for thinking and communication and for the way it sets up much of what comes after it.

3.2.2. Science, IBE, Confirmation

This CT course has been developed at a university (Simon Fraser University) that requires students, regardless of their field of study and major, to complete a minimum number of courses designated as Writing, Quantitative, and subject Breadth to ensure that every student's degree incorporates some minimum of basic competencies. This course qualifies as Quantitative, as well as Science (General) or Social Science breadth. Units 6-7 are the principal basis for the Quantitative designation, although Units 2-3 and 9 qualify in a looser sense. Units 4-5, 6, and 9-10 all support the Science and Social Science designations. I present Units 4 (Explanation and Confirmation) and 7 (Bayes′ Rule) in particular as topics in scientific reasoning.

The importance of CT is based in part on the importance of basic scientific literacy – for everyone, but especially for citizens of liberal democracies who have the ability and the responsibility to make important decisions of personal, local, national, and global significance. Specifically, it's based on the scientific aspects of fields of at least intermittent concern to everyone such as health and law, of areas of great importance for our collective future such as energy, environment, and economic development, and of various domains of inquiry, from cognitive science to cosmology, that contribute to our growing understanding of the place of humanity and the individual in the world. Full scientific literacy requires some functional awareness of the major knowledge products (discoveries, theories, and applications) of modern science, and of course by this standard quite a small fraction of society is scientifically literate. But the core of scientific literacy is a grasp of the basic methods of science, which includes the institutional and professional practices of science, and most fundamentally, what I'll call scientific reasoning, the formulation of hypotheses, explanatory or otherwise, and the use of evidence in confirming and disconfirming them. These are reasoning skills that enable people outside (and inside) of science to understand how it "works" and also to appreciate legitimate criticism and reject illegitimate criticism. By this standard, also, it appears that a dismayingly small fraction of society is scientifically literate, but there may be better prospects for improving it, in part through courses such as this one. As Peter Ellerton puts it, "making [scientific disciplines'] thinking explicit, and teaching students how to analyse and evaluate it, is key to developing scientific literacy, which must extend beyond simply knowledge of scientific content or even the understanding of scientific concepts" (2021: 9).

While there is a vast domain of technical and discipline-specific thinking that enables analysis and evaluation of science, scientific reasoning in its most general form, the form that it's feasible to incorporate into a CT course for students who typically have little or no science background, transcends not only specific scientific disciplines but also science in the academic sense. I allow everyday inquiry to count as scientific if it's undertaken in the right way. In *The Scientific Attitude*, Lee McIntyre asks whether everyday inquiries such as trying to determine where I left my keys should count as science if they exemplify the scientific attitude (caring about evidence and being willing to change theories on the basis of evidence) (2019: 73). Whether or not the full range of examples in this course constitute *science*, they can have the *character* of science — be *scientific* — insofar as they apply the reasoning methods of science. My approach, as Jonathan Haber puts it, "can be described as 'thinking like a scientist' but it would be more accurate to say that all critical thinkers, including scientists, rely on methods that, while inspired by the development of modern science, are relevant to every aspect of life" (2020: 16). In other words, I present CT as encompassing this aspect of science, its most central and general forms of reasoning.

I introduce scientific reasoning in Unit 4 (Explanation and Confirmation). This is the reasoning often labeled *abductive*, although I describe it, along with the reasoning of Units 8-10, as *inductive*. Charles Peirce's original discussion of abduction (CP 5.189) has the argument pattern:

- The surprising fact, *C*, is observed.
- But if *A* were true, *C* would be a matter of course.

There is reason to suspect that *A* is true.

Versions of this reasoning pattern long predate Peirce. Richard Jeffrey (1985: 56-57) notes that Descartes and Huygens described similar patterns. Peirce's abduction is distinctive because, as his conclusion indicates, technically he doesn't view the inference as supporting the truth of hypothesis A, but only as providing a reason to investigate A by other means. Characterizing this inference as inductive is convenient in a course with modest philosophical objectives, and it also suits my treatment of the argument, which I understand as raising the probability of the hypothesis in a way that can be described by Bayes' rule (see next section). I call the following pattern *confirmation*:

- If HYPOTHESIS then EVIDENCE (prediction).
- EVIDENCE (prediction true)

HYPOTHESIS is true.

Peirce elsewhere says that an abductive inference is justified by “its affording an explanation of the facts” (CP 1.139). This is one way I explain the conditional premise to students. “If HYPOTHESIS then EVIDENCE” links a hypothesis to a prediction, and we say the prediction is a good one (the conditional is acceptable) if we would deem the hypothesis to be a good explanation of this evidence, were it to occur. Notwithstanding Peirce’s connection of abduction with explanatory reasoning, the common equation of abduction with IBE is probably mistaken (Campos 2011; Mcauliffe 2015). Peircean abduction is “the process of forming an explanatory hypothesis” (CP 5.171) to investigate, reasoning that is what John Woods calls “ignorance-preserving yet action-motivating” (2021: 335). IBE is inductive, though it’s a “version of induction that is not ampliative, but rather strictly evaluative” (Mcauliffe 2015: 303). It’s a schema endorsed, albeit labeled as abduction, by Josephson and Josephson (1994) and Douglas Walton (2005), and labeled as induction by textbooks such as MacDonald and Vaughn (2019) (at least in places — see the discussion in section 3.1.2). My equivalent wording, highlighting the distinction between an explanation being best and it being true, is:

- EVIDENCE (observation)
- HYPOTHESIS is the <u>best</u> explanation of EVIDENCE.

HYPOTHESIS is the <u>true</u> explanation of EVIDENCE.

Since IBE is technically not abduction, and I present the Peircean pattern as confirmatory support rather than truly abductive, my “abduction” unit is not really on abduction, and it’s appropriate not to introduce the term to students. Putting the two types of arguments together in one unit, yet clearly distinguishing them from each other and treating them separately, allows me to capture the different language of many real or realistic examples. Some explicitly or implicitly invoke explanatory superiority in support of a hypothesis. Others use the language of prediction, such as “If HYPOTHESIS were true then we would see EVIDENCE, which we do see…”. I’m able to take these different cases on their own terms without teaching students that some of them must be transformed into a different pattern in order to be properly understood.

The IBE pattern with its "best explanation" premise serves as th e focus of a limited discussion of criteria of explanation, and as the intermediate conclusion in an extended argument whose sub-argument displays the use of these criteria and the essential process of considering alternatives. The confirmation pattern works well as part of a broader discussion of the role of background assumptions (auxiliary hypotheses) in disconfirmation, the focus of Assignment 1. Students learn that although the pattern is deductive, the false prediction doesn't result in conclusive disconfirmation of the hypothesis so long as a false background assumption is a live possibility. I believe this is an extremely important logical maneuver for students to learn. Much larger textbooks (MacDonald and Vaughn 2019) only briefly present the schema in which some observation is predicted jointly by a hypothesis and a background assumption or theory, and without taking students from that to the disjunctive conclusion that results from a false prediction. If these authors truncate their treatment of this reasoning because they regard it as something that would be intuitively obvious to most students, they are wrong. Only with reminders of De Morgan's laws and careful discussion of examples do most students see this. Finally, discussing the two patterns lays the groundwork for Unit 5 in which I discuss a family of errors under the loose category of *confirmation bias*. Some of these errors (ignoring alternative explanations) refer directly to IBE, others (imprecise prediction) refer directly to confirmation, and others (biased search for evidence and *ad hoc* explanation of disconfirming evidence) may be connected to reasoning in either form.

3.2.3. Probabilistic and Bayesian Reasoning

The third of the five questions I raised about this course in the introduction asks how it earns Simon Fraser University's Quantitative designation. The designation means that a course "will assist students to develop quantitative (numerical, geometric) or formal (deductive, probabilistic) reasoning" as well as general analytic skills. While any CT course that teaches any deductive reasoning could technically qualify as quantitative, such a minimal fulfillment of the criteria is not actually sufficient to keep the designation. The spirit of the Quantitative requirement is expressed in its purpose of "deepening the understanding and appreciation of quantitative and formal reasoning, their ubiquitous utility, and their creative potential". Having discussed deductive reasoning in section 3.2.1, in this section, I'll discuss how this course qualifies as quantitative through its numerical, geometric, and probabilistic components, including my presentation of the course's most complex quantitative thinking tool, Bayes' rule, and the relationship between the non-quantitative (IBE and confirmation) and quantitative (Bayesian belief updating) aspects of scientific reasoning.

Many students arrive in the course with weaker numeracy skills than one would expect based on their having graduated high school. Some are innumerate and some are math phobic. Components of the course in which students, for example, add and multiply simple fractions or interpret percentages serve to renew basic skills that are necessary for achieving the higher aims of the Quantitative designation. This is a low-level but not low-importance objective of the quantitative units. Some students are computationally competent in high-school or university-level mathematics but their quantitative reasoning skill is compartmentalized and mechanical; they struggle to grasp the meaning of a number in a simple word problem or the difference between saying that most drug dealers are murdered by gangs and saying that most gang murder victims are drug dealers. These students may be able to solve a quadratic equation but be totally flummoxed when asked why the chance of a coin landing both *Heads* and *Tails* on a single flip is, obviously, 0 rather than $^1/_4$ (as one would calculate with ***P***(*Heads*) × ***P***(*Tails*)). The purely mathematical ambitions of Unit 6 are modest. Through simple concrete examples we discover and use the probability rules (disjunction, negation, conjunction, weighted average for probability and expected value, at least one) and work out why they have the form they do. It appears there are few students for whom this process is superfluous. Time spent securing the foundations of probability, in particular conditional probability, is well spent.

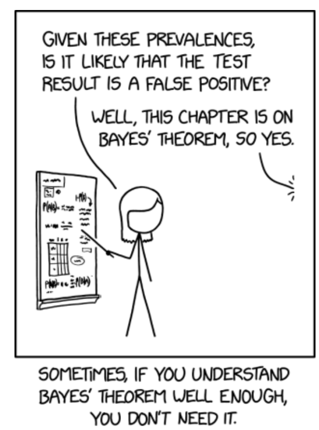
One very basic way in which the course achieves its quantitative goals is that it pushes students to use, think about, and express probabilities in a less compartmentalized way, not simply as a part of math problems disconnected from the rest of their reasoning. One way it does this is by having students work through examples in which people's degrees of belief are expressed in numerical terms (Unit 7), and having them provide and explain and use their own subjective probabilities (see Appendix B, Assignment 2). People's preference as receiver for the clarity of numerical probabilities, but as sender for the vagueness of verbal probabilities has been called the "preference paradox" (Erev and Cohen 1990) – interesting given how unstable, inconsistent, and tentative the deployment of probability vocabulary is in younger children (Groth *et al*. 2016). Although subsequent research has shown a greater preference for using numerical probabilities even for senders "when it really matters" (in high-stakes scenarios), perhaps based partly in people's awareness of the inconsistent use of a limited probability vocabulary (Dhami and Mandel 2022), this preference flip is not activated for most students in an academic context, and most are reluctant to give numerical subjective probabilities. Many sense that there is something illegitimate about subjective probabilities and need coxing to use probabilities in this way.

Of course the point of acknowledging the legitimacy of subjective probability is to actually use it. The physicist Sean Carroll quips that the way many people talk about their own beliefs suggests that they assign only probabilities of 1, 0.5 (i.e. indifference), or 0 to every proposition. It's an exaggeration but not a huge one. The concept of the probability interval 0-1 is a valuable thinking tool all on its own, even before the probability rules, because it is a way for people to think of, hold, and express degrees of belief that fall between dogmatism and pure agnosticism or skepticism or indifference. It is a mechanism of both epistemic humility and epistemic precision. Class discussion emphasizes, as Richard Jeffrey puts it, how "momentous" it is to assign probability 1 ("infinite odds") to a hypothesis: "you'd stake your all on its truth", thinking it "advantageous to stake your life upon it in exchange for any petty benefit" (1992: 1). (I do this with vivid examples involving pits of crocodiles.) Class exercises estimating probabilities for statements such as "SARS-CoV-2 jumped to humans in a wildlife market", "The 2040 Olympics will be held in India", etc., allow students to see that numerical expressions of probability can be meaningful and reasonable (or unreasonable) in such cases, even if they can't be straightforwardly calculated. Once probabilities are assigned, they can be treated like any other probabilities to be used with Bayes' rule.

Since Bayes' rule is an immediate consequence of the conjunction and disjunction rules, it's often treated as a mathematical triviality, even if a highly useful one. Very few first-year (or fourth-year) Humanities students perceive the rule, its derivation and expansion, or its use as trivial. At least upon initially seeing it, most think their only hope to succeed in the unit is to memorize it or be given a formula sheet. Nor is this redundant education even for the small number of students who have encountered it before, for example in an introductory statistics course. These students typically report that they at most simply memorized it as one of many revealed formulas without any sense of its special significance for rational belief change. The hurdle that Bayes' rule represents for many students in a CT course leads David Manley (2019) to try to teach Bayesian reasoning without ever actually showing or using the full standard form of the rule, which he calls "daunting". I'll discuss this approach below, but I devote a substantial unit of the course to constructing and using Bayes' rule because 1) I think the rule, while not something to which I overtly commit as a unifying principle of the whole course (see section 2.1.1), is the most important thinking tool for CT after the general concept of an argument; 2) It can be seen as the conversion of the useful odds form of probability into the familiar chance form (e.g. 2:1 = $^{2}/_{(2+1)}$), an exercise with which most students struggle at first but become more familiar by Unit 7; 3) I'm under no illusions that it's a trivial or spontaneous insight for most students; 4) I think that very little, and nothing lasting, is gained by having students memorize it or copy it from a supplied formula sheet; and 5) For students with even a modest commitment to working through it with me, it does not have to be daunting. My treatment of Bayes' rule thus differs from the approaches of many other textbooks, which ignore it even in a sprawling treatment of CT (e.g. MacDonald and Vaughn 2019; Hughes and Lavery 2008), cover it in an abrupt, dense, and non-visual manner that is useless to nearly all students who take this course (e.g. Salmon 2013), or use the alternative form of it (e.g. Manley 2019).

One way to see the value of Bayes' rule for CT is that it encapsulates multiple distinct principles for good reasoning about evidence. I teach these as four "Bayesian Lessons" that emerge throughout Unit 7.

1) Consider prior probabilities. Bayes' rule displays the prior probability, reminding us that it's impossible to determine where a piece of evidence gets us without an idea of our "starting point". Even though the repeated application of Bayes' rule allows for convergence from different starting points, there must *be* a starting point for any single use of evidence to guide a change of belief. Becoming fixated on the evidence and neglecting the prior probability or base rate is a recognized fallacy because it is, in fact, a very reliable error that I can count on an entire class making even after introducing prior probability and highlighting its importance. The startling importance of a base rate is a standard element of examples used to introduce Bayes' rule for good reason.



(Source: xkcd.com/2545. Licensed under Creative Commons Attribution-Noncommercial 2.5 License.)

"Starting point" is in scare quotes above because this way of talking about and using Bayes' rule imposes a kind of narrative on its operation. Despite the temporal connotations of *prior* and *posterior* (*updated*) that "suggest something diachronic", the rule is "entirely synchronic… on a par with all the other (synchronic) theorems of probability" (Hajek 2016: 8). *Prior* just means "without the evidence", not necessarily "before the evidence". Nonetheless, it is a useful narrative to get students to think of Bayes' rule as guiding a process of updating their (degree of) belief: having a degree of belief in some hypothesis based on everything they know so far, discovering new evidence, deciding how that new evidence should modify their degree of belief, and arriving at a new epistemic position that reflects both their old and their new evidence.

The concept of prior probability is so important for effective reasoning that it's tempting to build it into an approach to argument analysis (Hahn and Oaksford 2012). I acknowledge this in my account of the critical thinker (section 1.2.2) with the point under argument analysis about recognizing the "evidential burden" of an argument, based on the plausibility of its conclusion. But trying to formally Bayesianize argument analysis in an introduction to the topic risks making it overly complex. Hence I settle on the restrained pluralism approach, treating these as ultimately consistent but still separate modes of reasoning (though see below on how Bayesian reasoning relates to IBE).

2) <u>Confirming evidence: more likely if the hypothesis is true, less likely if the hypothesis is false</u>. This is fundamentally what we need to be able to say of any purported piece of supporting evidence. Most people will agree to this fairly readily when it's put to them as a general principle, but in practice it is all too easy to simply not fully consider how likely the evidence would be if the hypothesis is false. Bayes' rule not only presents both terms, it puts them in a fraction, showing evidence strength as a dimensionless quantity that ranges from zero to infinity. Less obviously, this also shows that the confirming strength of E may be very different from the disconfirming strength of ¬E. Noticing this requires an unusually solid grasp of the relevant fractions, but it's easy to see in the geometric presentation (Bayes box) I'll discuss below.

3) <u>If evidence E confirms then ¬E disconfirms</u>. This emerges out of a few quick exercises with the rule and is immediately clear in the Bayes box diagram. It gives a formal foundation to the intuition that confirmation bias is an error: it's not just intuitively an unfair response to evidence, it's mathematically nonsensical to accept that E counts in favour of a hypothesis without accepting that ¬E counts against it (though it may count against it to a different degree – see Lesson 2 above).

4) <u>Yesterday's updated probability is today's prior probability</u>. The large story-based exam questions of Unit 7 involve both hypothetical and iterative uses of Bayes' rule, the former for potential (within the story) evidence discoveries and the latter for subsequent evidence discoveries. The point of this latter type of question is for students to recognize updating as an ongoing process and Bayes' rule as the mechanism for rationally taking into account successive pieces of new information to adjust one's degree of belief. Lesson 4 is a lesson *for* the use of

Bayes' rule, but it's also a lesson that is, loosely speaking, "in" Bayes' rule – not directly in the mathematical form of the rule, as are Lessons 1-3, but in the narrative imposed on the rule by a literal sense of *prior* and *updated*/*posterior*. The lesson connects with the Unit 5 lesson about using our total evidence, but in a way that acknowledges that our total evidence is, after all, constantly changing and that carefully *building* on yesterday's rational belief is equivalent to, and usually more natural than, starting over every day with an ever-expanding set of total evidence.

There is still the question of what formal version of Bayes' rule to use, and the concern that the full rule is overwhelming for students in a first-year Humanities course. The allegedly less daunting way to perform a Bayesian update goes like this (e.g. Manley 2019; O'Hagan 2006):

1) Represent the prior probability as an odds ratio.
2) Determine the evidence strength (Bayes' factor): $^{\boldsymbol{P}(E|H)}/_{\boldsymbol{P}(E|\neg H)}$.
3) Find the updated (posterior) odds by multiplying the prior odds by the evidence strength.
4) As desired, convert the updated odds into the chance form to express it as a fraction, percentage, etc.: A:B = $^{A}/_{(A + B)}$. (Or: $^{odds}/_{(1 + odds)}$.)

For example:

1) We think that hypothesis H has a prior probability of $^1/_3$. We represent that as the odds ($^1/_3$):(1 – $^1/_3$) or 1:2. (This may be expressed as the odds ratio 0.5, signifying that H is half as likely as ¬H.)
2) We think that evidence E has strength 3 because we've estimated that $\boldsymbol{P}(E|H) = 0.45$ and $\boldsymbol{P}(E|\neg H) = 0.15$.
3) For the updated odds, we multiply the prior odds by the evidence strength: 1:2 × ($^{0.45}/_{0.15}$) = 0.45:0.3 = 1.5. (Or we multiply 0.5 × 3 = 1.5.)
4) Optionally we convert the updated odds into a chance probability: $^{0.45}/_{(0.45 + 0.3)}$ = $^{0.45}/_{0.75}$ = 0.6. (Or $^{1.5}/_{(1 + 1.5)}$ = 0.6.)

The method has the supposed advantage of not sparking panic among students by never confronting them with Bayes' rule in its entirety. The method makes most sense if one can fully commit to working with odds, omitting step 4. But since so much probability information is in the form of fractions and percentages, this is not a good policy, especially if the method represents an odds ratio not like the score in a sports match, the analogy I use to get most students to understand it, but as a single number such as 0.5, which students will confuse with a chance of 0.5. (Many struggle to distinguish chance and odds even when they're presented in completely different formats: $^1/_2$ vs. 1:2.) Producing a final answer with a fraction whose denominator is (1 + odds) is guaranteed to be met with puzzled responses asking where the 1 “comes from” – it looks sort of like the “at least one” solutions where we subtract a probability from 1, but isn’t that. And it passes up a sort of example that is easy and useful as a stepping stone to Bayes’ rule, cases like the fire alarm (see Appendix F). With minimal prompting, students can usually articulate why they believe there is a low chance of fire even when the fire alarm rings: the value of the fraction (Alarms with fire) / (Total alarms) is low. But the reason it’s low is because the denominator (Total alarms) = (Alarms with fire) + (Alarms without fire), and (Alarms without fire) is so much greater than (Alarms with fire). This example, which lays out the basic fraction structure of Bayes’ rule, is intuitive to almost everyone. Most readily accept it as an explanation of why the chance of fire is low despite the alarm. This sort of example sets up a presentation of the full rule. Once students have this example, it makes sense and works well to push on to the full rule.

I construct the full form of Bayes’ rule on the foundation of the conjunction rule, to which I devote considerable time and numerous examples in Unit 6 so that students understand why we are multiplying and why we are conditionalizing, as well as the fire alarm case. Nothing in my construction of the rule is presented as a “just do this” step. Manley’s approach does have the advantage of separately presenting the two key pieces of information needed for a Bayesian update – the prior probability and the evidence strength – but the question of how to connect them gets an opaque answer. While it may be simple and easy enough to *remember* “multiply the prior odds by the evidence strength”, this is just a stipulation not grounded in the conjunction rule or a geometrical representation (see below). (Multiplication is a thing you can do with two numbers, so let’s multiply them!) The course manual, the slides, the widget, and the examples and problems all separately call attention to the two key numbers, but rather that telling students “just multiply them”, I show how they naturally find places in an elegant rule whose form can be fully explained with a simple rearrangement of the conjunction rule, everyday examples like the fire alarm, and toy cases worked through with the geometric method in which multiplication represents the transformation of a one-dimensional line segment into a two-dimensional rectangle.

I developed the geometric representation of a Bayesian odds update from a simple illustration to something more integral to the explanation and the reasoning process, the “Bayes box” (Unit 7.i).[23] The diagram achieves several things:

[23] I began teaching a version of the Bayes box around 2014. I developed it for the course manual, the class slides, the exam and assignment, and the web widget over the following years. It was with a mix of

- My initial presentation can use icon array graphics as an initial step in constructing the purely geometric diagram, by configuring icon arrays and then overlaying the simple geometric graphic (see Appendix F). Icon arrays are a proven effective way to communicate risk and probability information, particularly in medical contexts, in part because they offer a very natural way to grasp part-to-whole relationships (Ancker 2006).

- It reinforces that three pairs of values (***P***(H) and ***P***(¬H), ***P***(E|H) and ***P***(¬E|H), and ***P***(E|¬H) and ***P***(¬E|¬H)) each add to 1 by showing each as a divided line segment.

- It reinforces that the probabilities ***P***(E|H) and ***P***(E|¬H) are independent (most students believe at first that these must add to 1 even after seeing examples where they don't) by showing them as independently set features of the box.

- It simultaneously shows the odds updated on E and the odds updated on ¬E in comparisons of either the two shaded or the two empty rectangles.

- It reinforces the relationship between the odds and chance forms with the simple question "What percentage of the total shading is in the Hypothesis rectangle (green in the slides and widget)?".

- It reinforces why we are multiplying probabilities, since nearly all students understand that the area of a rectangle is the product of its base times its height.

- It presents conditional probabilities in concrete form with restricted regions within the box. For the evidence likelihoods: conditionalizing on H means looking at the H (left) side; conditionalizing on ¬H means looking at the ¬H (right) side. For the updated H probabilities: conditionalizing on E means looking at the total shading; conditionalizing on ¬E means looking at the total empty space.

---

gratification and dismay that I recently discovered that the popular YouTube mathematics channel 3Blue1Brown released an episode in 2020 on "Bayes theorem, the geometry of changing belief" with the same geometric explanation. As of 2023 January, it had approximately 3 million views and thousands of enthusiastic comments commending the graphical approach as vivid, intuitive, and superior to standard presentations. I teach Bayes' rule this way precisely because I believe it is intuitive with the right prompts – it is the obvious way to graphically represent a Bayesian odds update – so this does not surprise me. If an excellent creator of graphical explanations of mathematical concepts *didn't* do something like this in an episode devoted to Bayes' rule, I would be concerned that it's less intuitive than I think it is. So far I've not seen it used in any other CT textbooks.

- It allows students to draw on perceptions and intuitions of size and mass to cognize the odds, the “balance of probabilities”, as a concrete reality.

My initial use of the graphic added to the presentation of Bayes’ rule as just a neat way to illustrate it. It came across to students as an afterthought or a piece of extra work and did not have very encouraging results – it was difficult to get them to see that understanding the diagram *is* understanding the rule. This began to change as I began integrating the diagram into the presentation of the rule rather than after the fact, and as I developed a dynamic and interactive web version of the diagram to supplement the more or less static diagrams (the slides are extensively animated) and drawing exercises. Now more students, including some with low mathematical confidence (and ability), report that the diagram actually helps them, and even that it makes a difficult section of the course a bit fun. I designed the web widget, with coding by a former Douglas College student, Anderson Chen. The widget can be accessed at Chen's Github: **achenny.github.io/bayes-box-webpage**

The widget includes a Bayes’ rule calculator that is graphically tied to an interactive Bayes box. The calculator drives the graphic and the graphic drives the calculator. A user can move the sliders, for example watching how the updated probability (the positive predictive value of a test) changes with ***P***(E|¬H) (the false positive rate), seeing how evidence strength is independent of prior probability, etc. Evidence strength is highlighted as a separate value, and a toggle switches the operation from ***P***(H|E) mode to ***P***(H|¬E) mode for updating on different possible observations (prediction true or prediction false, if E is thought of as a prediction). Finally, the widget can operate as a sort of game for students who are unsure if they are attending to the relevant aspects of the graphic and who may still be unsure about converting odds to chance probability. A toggle switches the operation to visual-only mode and a button randomizes the values. Students can estimate the updated probability visually and then switch back to calculator mode to check their exact value.

The value of the Bayes box as an instructional aid, as a component of exercises and questions that is constructed or investigated by students, and as an interactive web tool is not based on any assumptions about the existence of “visual learners”. It’s not designed to cater to some students’ alleged visual “learning style”. While belief in learning styles is widespread and persistent, the theory lacks empirical support (Norman 2009). Nor is the Bayes box based on a faith that clear diagrams are automatically processed by learners in the expected way, which they often are not (Kottmeyer *et al*. 2020). The activity of drawing or manipulating the diagram is intended to align with a trend in science education that increasingly recognizes that “creating and revising visual representations can lead to deep understanding of the scientific concepts”, a process Tippett (2016) calls “learning with, rather than learning from, visual representations”. More specifically, it’s based on the idea that well constructed abstract diagrams can offer a learning advantage through “using space to think” by the “offloading of cognitive processes onto automatic perceptual processes” (Hegarty and Stull 2012).

In classroom exercises, students often make simple calculation errors that produce absurd answers, but then simply accept the number displayed by their calculator. Forcing students to first sketch a Bayes box and get a visual sense of what answers make sense is a way for them to catch careless calculation errors as well as learn the operation in a geometric way. The fact that students readily accept obviously wrong answers produced by their calculators highlights the superficiality of much quantitative learning and problem solving, and the need for students to think about what they are doing in some way other than mechanical calculation. The diagrams are a tool by which students can demonstrate, not just advance, their understanding of Bayesian updating. One reason the unit is challenging for even some mathematically adept students it that it’s designed, as much as possible in a course of this nature, to demand more than superficial, mechanical competence. A student who can compute an answer with Bayes’ rule but cannot interpret, explain, or produce the diagram does not understand what they’re doing with the rule with the same depth as one who can.

Several influential definitions of CT, and my own, acknowledge decision-making as an important element of CT. CT textbooks cover decision-making in a variety of ways, not always quantitative. For example, Woods *et al*. (2000) completely ignore it; Hughes and Lavery (2008) and MacDonald and Vaughn (2019) have brief non-quantitative sections on ethical reasoning; Salmon (2013) runs through some standard decision theory in the same student un-friendly style as the rest of the text; Manley (2019) has engaging sections on basic decision theory and some findings from behavioural economics (e.g. the well known framing effects), though his suggested assessments are dispiriting: mostly multiple-choice questions about the words that Manley writes about decision-making. Of course covering decision-making in any depth carries an opportunity cost in the form of other important topics not covered, and I set this trade-off differently than Manley does. Unit 6 includes some basic expected value problems which connect well with the treatment of Bayes’ rule by having students work with examples of the “sum of products” structure (for a probability-weighted average value) they will see in the next unit in the denominator of Bayes’ rule. (Manley constructs his treatment of Bayes’ rule partly to shield students from this math but then uses it anyway in his section on expected value – apparently students can handle this, after all.) Unit 7 incorporates a simple and natural way to make decisions an aspect of problems of belief change. This is in contrast to decision theory problems where all the information is laid out and the question is whether to do A or B. Here the student works through a multi-step narrative problem in which a character has reflected and determined a probability threshold at which they will act in a particular way, based on personal considerations of value and risk. The question of how they should act becomes an extension of how they should update their beliefs. These questions help fill out the large multi-part Bayes' rule problems into more complete scenarios of belief and decision. It's a simple way to show students for whom quantitative operations on their own beliefs are an intensely foreign way of thinking, and who may not yet be intrinsically motivated to always hold the most coherent and evidence-responsive beliefs they can, another way in which this type of belief-updating procedure could matter for living.

I'll conclude by looking at how the Unit 7 treatment of Bayesian reasoning relates to the Unit 4 treatment of IBE and confirmation. These units employ the same terms of *hypothesis* and *evidence*. How should a student understand the relationship between these topics? Are they simply different thinking tools to be used in different sorts of discussions? To some extent, the answer is "yes". This is part of the restrained pluralism of the course's approach to CT. But since I unite the topics in a common project called *scientific reasoning* and use some common terms in doing so, they should have a basic coherence. While it would not be helpful to students working to understand how Unit 7 is grounded in Unit 6 for me to repeatedly highlight its connections to Unit 4 in an effort to show that they "match up", I do want the connections to be there. The greatest concern about teaching both IBE and Bayesian reasoning would be if they were outright incompatible, such that teaching both to students is teaching ways of reasoning that implicitly contradict each other. This incompatibilist thesis is argued by Bas van Fraassen (1989), who claims that IBE adds a kind of probability bonus, not recognized in the terms of Bayes' rule, to the best explanation (and so much the worse for IBE). As in the free will literature, incompatibilism has not won many adherents, but compatibilists disagree among themselves. Generic compatibilism is the view that "the Bayesian either does or should make use of explanatory considerations in assigning probabilities (priors and/or likelihoods) to the hypotheses" (Henderson 2014: 688) with no explanatory probability bonus.

The Unit 4 treatment of IBE identifies two criteria of explanatory adequacy: simplicity and conservativism, which I call "fit with background information". These are what the compatibilists Zachary Wojtowicz and Simon DeDeo (2020) call "theoretical values", the two components of prior probability, in their decomposition of Bayes' rule into explanatory values. Simplicity encompasses parsimony, concision, and unification (which can vary independently, making judgements of simplicity difficult in practice), although students need only recognize general appeals to simplicity in common language, not differentiate these aspects of simplicity, for the examples and questions in this part of the unit. An example we look at in class from an editorial in *Nature* points to the "convoluted scenarios" one must entertain to explain COVID-19 superspreading events by contaminated surfaces (instead of airborne transmission) (Lewis 2021). We talk about what makes one hypothesis simpler by talking through the scenarios but not by trying to separately score them for concision, parsimony, and unification. Although they can vary independently, in many useful short examples, they align. The convoluted surface contamination scenarios have less concise descriptions, they are less parsimonious (requiring more elements or causal links), and surface contamination may be a less unifying explanation, for example if we needed to invoke airborne transmission anyway to explain other cases of transmission. Wojtowicz and DeDeo define what I call "fit with background information" as the "influence of domain-specific tacit knowledge and background knowledge (such as base rates)" (2020: 982). Although the *Nature* editorial doesn't explicitly advance a fit consideration, the IBE argument could be expanded with reference to a base rate in a domain such as *respiratory viruses*.

Unit 4 doesn't introduce separate criteria of explanatory adequacy that are assigned to Wojtowicz and DeDeo's "empirical values", those which contribute to the Bayesian likelihood ($\boldsymbol{P}$(E|H)). They distinguish the probability of the individual observed facts in E, based on H, and the extent to which H makes probable the overall pattern or collection of observed facts in E. When students consider and explain the priors they assign to competing hypotheses in Assignment 2, they are to think about the criteria from Unit 4. Consideration of evidence strength is therefore a new task in Unit 7. In Unit 4, the likelihoods ($\boldsymbol{P}$(E|H)) of competing explanations are implicitly assumed to be high. Depending on how a hypothesis is formulated, it can even make sense to think of the likelihood as 1. (For example, if the hypothesis is "X caused E", so the appearance of E is "built in" to the hypothesis, then $\boldsymbol{P}$(E|H) = 1.) Then hypotheses are effectively being compared by their priors. Unit 5 inches closer to the Bayesian perspective in its discussion of, for example, ignoring alternative explanations, which is Bayesian in all but name: an alternative explanation H2 is a way in which the evidence could occur even if hypothesis H1 is false — it boosts $\boldsymbol{P}$(E|¬H1), making E weaker evidence for H1. One can look at the imprecise predictions of Unit 5 in a similar way. Being less precise makes the prediction E count as true in a wider range of scenarios, including more scenarios in which H is false, so this boosts $\boldsymbol{P}$(E|¬H1).

This means that the model of IBE is somewhat restricted compared to the treatment of Bayes' rule, partly in the interests of efficiency and because students are here seeing the first extended arguments, which are already much more complex than the syllogisms of Units 2-3. This is not an incompatibility because the potential incompatibility arises from an *extra* probability component recognized in IBE. What would threaten incompatibilism is following the example of a textbook like MacDonald and Vaughn (2019), with a list of criteria of explanatory adequacy that places testability and fruitfulness alongside simplicity and conservativism. My Unit 4 makes clear that testability is important for methodological reasons. A hypothesis must make a checkable prediction in order to be confirmable, and there may be reasons to treat an unconfirmable hypothesis differently from a confirmable one. But testability *per se* does not earn it a probability bonus. MacDonald and Vaughn fudge this by saying that a testable theory is "superior" (2019: 363), without directly claiming that it's thereby more probable – but also without denying this. A student will have even more difficulty determining whether *superior* is to be understood in probabilistic terms when they turn to fruitfulness (making successful novel predictions (Kuhn 1977: 322)), which MacDonald and Vaughn say makes a theory both superior and more likely to be true, which suggests that "superior" testability does confer a probability bonus. Whether fruitfulness makes a theory more probable at all is disputed (e.g. Lacey 1999: 60). The compatibilist Leah Henderson (2014) allows that it does, though Wojtowicz and DeDeo find no explicit place for it in their decomposition of Bayes' rule. If fruitfulness is to be accommodated within a Bayesian framework, we must be able to say that the likelihood of some observed evidence, given a hypothesis that specifically predicts it, is higher than it is, given a hypothesis that is simply silent about that evidence (neither predicts it nor predicts contrary evidence). Otherwise it constitutes some kind of probability bonus. Maybe we can say that, but it seems wiser, in the interests of a compatibility that is reasonably transparent to students in a CT course, to focus on cases where hypotheses compete to explain the same set of facts, whether already observed or predicted and then observed. So although the course does not explicitly present a fully worked out IBE-Bayes compatibilism within its treatment of scientific reasoning, adequate connections are there, especially for the curious and attentive student. And although students learn Bayes' rule as the more powerful, more encompassing thinking tool, and it's the tool that they do more with in the exam and the assignment, this doesn't imply a rejection or replacement of IBE. These are different tools, and students learn how to use and participate in discussions with both.

References

3Blue1Brown. (2020). Bayes theorem, the geometry of changing beliefs [Video]. YouTube. www.youtube.com/watch?v=HZGCoVF3YvM.

Abrami, P. et al. (2015). Strategies for teaching students to think critically: A meta-analysis. *Review of Educational Research* **85**(2), 275-314.

Ancker, J. S. *et al*. (2006). Design features of graphs in health risk communication: A systematic review. Journal of the American Medical Informatics Association **13**(6), 608-618.

Anderson, L. *et al*. (2001). *A Taxonomy for Learning, Teaching and Assessing: A Revision of Bloom′s Taxonomy of Educational Objectives*, New York: Longman.

Arnauld, A. and Nicole, P. (1996 [1662]). *Logic or the Art of Thinking* (J. V. Buroker (trans.)). Cambridge: Cambridge University Press.

Arum, R. and Roksa, J. (2011). *Academically Adrift*: *Limited Learning on College Campuses*. Chicago: University of Chicago Press.

Atkinson, D. (1997). A critical approach to critical thinking in TESOL. *TESOL Quarterly* **31**(1), 71-94.

Bachiochi, E. (2022 July 1). What makes a fetus a person? The New York Times. www.nytimes.com/2022/07/01/opinion/fetal-personhood-constitution.html

Bailing, S. (1995). Is critical thinking biased? Clarifications and implications. *Educational Theory* **45**(2), 191-197.

Bailin, S. *et al*. (1999a). Common misconceptions of critical thinking. *Journal of Curriculum Studies* **31**(3), 269-283.

Bailin, S. *et al*. (1999b). Conceptualizing critical thinking. *Journal of Curriculum Studies* **31**(3), 285-302.

Barron, A. and Klein, C. (2016). What insects can tell us about the origin of consciousness. *PNAS* **113** (18), 4900-4908.

Barta, P. (2020). Norton′s material theory of analogy. *Studies in History and Philosophy of Science Part A* **82**, 104-113.

Battersby, M. (2006). Applied epistemology and argumentation in epidemiology. *Informal Logic* **26**(1), 41-62.

BC′s Curriculum. (2022). curriculum.gov.bc.ca

Bentham, J. (1824). *The Book of Fallacies*. In Schofield, P. *The Collected Works of Jeremy Bentham*: *Political Writings*. Oxford: Clarendon Press.

Blair, A. (2021). Arguments and Critical Thinking. In Blair, A. (ed.) *Studies in Critical Thinking, 2nd Ed*. Windsor, ON: Windsor Studies in Argumentation.

Bloom, B. *et al*. (1956). *Taxonomy of Educational Objectives. Handbook I*: Cognitive Domain, New York: McKay.

Brookfield, S. (2011). Te*aching for Critical Thinking*: *Tools and Techniques to Help Students Question their Assumptions*. San Fransisco: Jossey-Bass.

Cahill, A. and Bloch-Schulman, S. (2012). Argumentation step-by-step: Learning critical thinking through deliberate practice. *Teaching Philosophy* **35**(1), 41-62.

Campos, D. (2011). On the distinction between Peirce′s abduction and Lipton′s inference to the best explanation. *Synthese* **180**(3), 419-442.

Capossela, T. 1998. What is critical writing? In Capossela, T. (ed). *The critical writing workshop: Designing writing assignments to foster critical thinking*. Portsmouth, NH: Heinemann.

Case, R. (2013). The unfortunate consequences of Bloom′s Taxonomy. *Social Education* **4**, 196-200.

Davies, W. M. (2006). An ‶infusion″ approach to critical thinking: Moore on the critical thinking debate. *Higher Education Research & Development* **25**(2), 179-193.

Davies, M. *et al*. (2021). Using computer-aided argument mapping to teach reasoning. In Blair, A. *Studies in Critical Thinking, 2nd Ed*. Windsor, ON: Windsor Studies in Argumentation.

DeBoer, G. (2000). Scientific literacy: Another look at its historical and contemporary meanings and its relationship to science education reform. *Journal of Research in Science Teaching* **37**(6), 582-601.

Dewey, J. (1910). *How We Think*, Boston: D.C. Heath.

Dhami, M. and Mandel, D. (2022). Communicating uncertainty using words and numbers. *Trends in Cognitive Sciences* **26**(6), 514-526.

Doya, K. *et al*. (2006). *Bayesian Brain: Probabilistic Approaches to Neural Coding*. Cambridge, MA: MIT Press.

Duncan, W. (1748). *The Elements of Logic*. New York: Nichols & Co.

Ellerton, P. (2021). Critical thinking and the methodology of science. Presentation to the Australasian Science Education Research Association.

Ellerton, P. (2022). On critical thinking and content knowledge: A critique of the assumptions of cognitive load theory. *Thinking Skills and Creativity* **43**, 1-7.

Ellerton, P. and Kelly, R. (2022). Creativity and critical thinking. In Berry, A. *et al*. (eds.) *Education in the 21st Century: STEM, Creativity and Critical Thinking*. Springer, 9-27.

Engber, D. (2021 Nov 24). The lab-leak theory meets its perfect match. *The Atlantic*. www.theatlantic.com/ideas/archive/2021/11/lab-leak-covid-origin-coincidence-wet-market/620794

Ennis, R. (1962). A concept of critical thinking: A proposed basis for research on the teaching and evaluation of critical thinking Ability. *Harvard Educational Review* **32**(1), 81−111.

Ennis, R. (1964). A definition of critical thinking. *The Reading Teacher* **17**(8), 599-612.

Ennis, R. (1987). A taxonomy of critical thinking dispositions and abilities. In Boykoff, J. and Sternberg, R. (eds.), *Teaching Thinking Skills: Theory and Practice*, New York: W. H. Freeman, 9−26.

Ennis, R. (1989). Critical thinking and subject specificity. *Educational Researcher*, **18**(3), 4-10.

Ennis, R. (1991). Critical thinking: A streamlined conception. *Teaching Philosophy* **14**(1), 5-24.

Ennis, R. (2013). Critical thinking across the curriculum. OSSA Conference Archive. 44. scholar.uwindsor.ca/ossaarchive/OSSA10/papersandcommentaries/44.

Erev, I. and Cohen, B.L. (1990) Verbal versus numerical probabilities: efficiency, biases, and the preference paradox. *Organizational Behavior and Human Decision Processes* **45**(1), 1–18.

Facione, Peter A. (1990). *Critical Thinking*: *A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction*, Research Findings and Recommendations Prepared for the Committee on Pre-College Philosophy of the American Philosophical Association, ERIC Document ED315423.

Finnocchiaro, M. (1981). Fallacies and the evaluation of reasoning. *American Philosophical Quarterly* 18, 13−22.

Finnocchiaro, M. (1997). The *Port-Royal Logic*′s theory of argument. *Argumentation* **11**, 393-410.

Floyd, C. (2011). Critical thinking in a second language. *Higher Education Research and Development* **30**(3), 289-302.

Gabbay, D. and Woods, J. (eds). (2008). *Handbook of the History of Logic* Vol.4. Elsevier.

Galef, J. (2021). *The Scout Mindset*: *Why Some People See Things Clearly and Others Don′t*. Portfolio.

Geach, P. (1979). On teaching logic. *Philosophy* **54**(207), 5-17.

Gentner, D. *et al*. (2001). Metaphor is like analogy. In Gentner, D. *et al*. *The Analogical Mind: Perspectives from Cognitive Science*. Cambridge, MA: MIT Press

Glaser, R. (1941). *An Experiment in the Development of Critical Thinking.* New York: Bureau of Publications, Teachers College, Columbia University.

Glaser, R. (1984). Education and thinking: The role of knowledge. *American Psychologist* **39**, 93-104.

Goddu, G. (2021). Logic and critical thinking. In Blair, A. *Studies in Critical Thinking, 2nd Ed*. Windsor, ON: Windsor Studies in Argumentation.

Govier, T. (1987). *Problems in Argument Analysis and Evaluation*. Dordrecht: Foris.

Govier, T. (1989). Analogies and missing premises. *Informal Logic* **11**(3), 141-152.

Govier, T. (1999). *The Philosophy of Argument*. Newport News, VA: Vale Press.

Govier, T. (2021). Reflections on minimal adversariality. *Informal Logic* **41**(4), 523-537.

Groth, R. *et al*. (2016). Overcoming challenges in learning probability vocabulary. *Teaching Statistics* **38**(3), 102-107.

Guthrie, J. and Springer, M. (2004). A Nation at Risk revisited: Did "wrong" reasoning result in "right" results? At what cost? *Peabody Journal of Education* **79**(1), 7-35.

Haber, J. (2020). *Critical Thinking*. Cambridge, MA: MIT Press.

Hahn, U. and Oaksford, M. (2012). Rational Argument. In Holyoak and Morrison (eds). *The Oxford Handbook and Thinking and Reasoning*. Oxford: Oxford University Press, 277-298.

Haidt, J. (2021 Nov 21). The dangerous experiment on teen girls. *The Atlantic*. www.theatlantic.com/ideas/archive/2021/11/facebooks-dangerous-experiment-teen-girls/620767.

Hayes, D. (2015). Against critical thinking Pedagogy. *Arts & Humanities in Higher Education* **14**(4), 318-328.

Henderson, L. (2014). Bayesianism and inference to the best explanation. *British Journal of Philosophy of Science* **65**, 687-715.

Hegarty, M. and Stull, A. (2012). Visuospatial thinking. In Holyoak and Morrison (eds). *The Oxford Handbook and Thinking and Reasoning*. Oxford: Oxford University Press, 606-631.

Henderson, L. (2014). Bayesianism and inference to the best explanation. *British Journal of Philosophy of Science* **65**, 687-715.

Hendrick, C. (2016 Dec 5). Why schools should not teach general critical-thinking skills. *Aeon*. aeon.co/ideas/why-schools-should-not-teach-general-critical-thinking-skills.

Hitchcock, D. (2017). *On reasoning and argument*: *Essays in informal logic and critical thinking*. Springer.

Hughs, W. and Lavery, J. (2008). *Critical Thinking*: *An Introduction to the Basic Skills*, *5th ed*. Broadview Press.

Hundleby, C. (2019). Aggression, politeness, and abstract adversaries. In Puppo, F. (ed.) *Informal Logic: A 'Canadian' Approach to Argument*. Windsor Studies in Argumentation.

Ikuenobe, P. (2001). Teaching and assessing critical thinking abilitiesas outcomes in an informal logic course. *Teaching in Higher Education* **6**(1), 19-32.

Jaschik, S. (2012 June 29). Texas GOP vs. critical thinking. *Inside Higher Ed*. www.insidehighered.com/quicktakes/2012/06/29/texas-gop-vs-critical-thinking.

Jeffrey, R. (1985). *Probability and the Art of Judgment*. Cambridge: Cambridge University Press.

Johnson, R. and Blair, A. (2006). *Logical Self-Defense*. New York: International Debate Education Association.

Johnson-Laird, P. (2010). Mental models and human reasoning. *PNAS* **107**(43), 18243-18250.

Josephson, J. and Josephson, S. (Eds.) (1994). *Abductive Inference*: *Computation, Philosophy, Technology*. Cambridge: Cambridge University Press.

Kenyon, T. and Beaulac, G. (2014). Critical thinking education and debiasing. *Informal Logic* **34**(4), 341-363.

Kottmeyer, A. *et al*. (2020). Diagram comprehension ability of college students in an introductory biology course. *Advances in Physiological Education* **44**, 169-180.

Kuhn, T. (1977). Objectivity, value judgment, and theory choice. In Kuhn, T. *The Essential Tension*. Chicago: University of Chicago Press: 320–39.

Laks, A. and Most, G. (2016). *Early Greek Philosophy Volume V*. Cambridge, MA: Harvard University Press.

Lacey, H. (1999). *Is Science Value Free? Values and Scientific Understanding*. London: Routledge.

Lederman, D. (2013 May 13). Less academically adrift? *Inside Higher Ed*. www.insidehighered.com/news/2013/05/20/studies-challenge-findings-academically-adrift

Lewis, D. (2021 Jan 29). COVID-19 rarely spreads through surfaces. So why are we still deep cleaning? *Nature News*, https://www.nature.com/articles/d41586-021-00251-4

Linardon, J. *et al*. (2017). The efficacy of cognitive-bahavioral therapy for eating disorders: A systematic review and meta-analysis. *Journal of Consulting and Clinical Psychology* **85**(11), 1080-1094.

Lipman, M. (1988). Critical thinking: What can it be? *Analytic Teaching* **8**(1), 5-12.

Lipton, P. (2004). *Inference to the Best Explanation, 2nd Ed*. London: Routledge.

Locke, J. (1975 [1690]). *An Essay Concerning Human Understanding*. Nidditch, P. (ed.) Oxford: Clarendon Press.

MacDonald, C. and Vaughn, L. (2019). *The Power of Critical Thinking, 5th Canadian Edition*. Don Mills, ON: Oxford University Press Canada.

Manley, D. (2019). *Reason Better*: *An Interdisciplinary Guide to Critical Thinking*, version 1.4. Tophat.

Marshall, J. D. (2001). A critical theory of the self: Wittgenstein, Nietzsche, Foucault. *Studies in Philosophy and Education* **20**, 75−91.

Maxmen, A. (2022 Feb 27). Wuhan market was epicentre of pandemic′s start, studies suggest. *Nature News*. www.nature.com/articles/d41586-022-00584-8.

Mcauliffe, W. (2015). How did abduction get confused with inference to the best explanation? *Transactions of the Charles S. Peirce Society* **51**(3), 300-319.

McIntyre, L (2020). *The Scientific Attitude*: *Defending Science from Denial, Fraud, and Pseudoscience*. Cambridge: MIT Press.

McKerrow, R. (1987). Richard Whately and the revival of logic in nineteenth century England. *Rhetorica* **5**(2), 163-185.

McPeck, J. (1981). *Critical Thinking and Education*. Toronto: Oxford University Press.

McPeck, J. (1990). Critical thinking and subject specificity: A reply to Ennis. *Educational Researcher* **19**(4), 10-12.

Mercier, H. and Sperber, D. Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences* **34**, 57-111.

Mill, J. S. (1843). *System of Logic*. In Robson, J. M. *The Collected Works of John Stuart Mill*. Toronto: University of Toronto Press.

Mill, J. S. (1859). *On Liberty*. In Robson, J. M. *The Collected Works of John Stuart Mill*. Toronto: University of Toronto Press.

Moulton, J. (1983). A Paradigm of Philosophy: The Adversary Method. In: Harding, S and Hintikka, M. (Eds.), *Discovering Reality: Feminist Perspectives on Epistemology, Metaphysics, Methodology, and Philosophy of Science*. Dordrecht, Holland: D. Reidel, 149-164.

National Commission on Excellence in Education. (1983). *A nation at risk*: *The imperative for educational reform*. Washington, DC: US Government Printing Office.

Norman, G. (2009). When will learning style go out of style? *Advances in Health Sciences Education* **14**, 1-4.

Norton, J. (2018). *The material theory of induction*. Posted version of June 26, 2018. https://www.pitt.edu/~jdnorton/papers/material_theory/material_theory_June_26_2018_toc.pdf.

O'Hagan, T. (2006 December). Bayes factors. *Significance*, 184-186.

Owen, S. (2012 Aug 11). Gail Collins says Texas GOP platform calls for schools to stop teaching "critical thinking". *Politifact*. www.politifact.com/factchecks/2012/aug/11/gail-collins/gail-collins-says-texas-gop-platform-calls-schools.

Paul, R. (1985). Bloom's taxonomy and critical thinking instruction. *Educational Leadership* **42**(8), 36-39.

Paul, R. (1992). *Critical Thinking*: *What Every Person Needs to Survive in a Rapidly Changing World*. Santa Rosa, CA: Foundation for Critical Thinking.

Peirce, C. S. *Collected Papers of Charles Sanders Peirce* (CP), 8 Volumes. Hartshorne, C. et al. (Eds.) Cambridge, MA: Harvard University Press.

Penkauskienė, D. et al. (2019). How is critical thinking valued by the labour market? Employer perspectives from different European countries. *Studies in Higher Education* **44**(5), 804-815.

Perkins, D. *et al*. (1993). Beyond abilities: A dispositional theory of thinking. *Merrill-Palmer Quarterly* **39**(1), 1-21.

Pinto, R. (2019). Argumentation and the force of reasons. In Puppo, F. (ed.) *Informal Logic*: *A 'Canadian' Approach to Argument*. Windsor Studies in Argumentation.

Possin, K. (2014). Critique of the *Watson-Glaser Critical Thinking Appraisal Test*: The more you know, the lower your score. *Informal Logic* **34**(4), 393-416.

Puppo, F. (2019). *Informal Logic: A 'Canadian' Approach to Argument*. Windsor Studies in Argumentation.

Reid, M. *et al*. (2007). Skill acquisition in tennis: Research and current practice. *Journal of Science and Medicine in Sport* **10**, 1-10.

Robinson, S. V. (2011). Teaching logic and teaching critical thinking: Revisiting McPeck. *Higher Education Research & Development* **30**(3), 275-287.

Rooney, P. (2010). Philosophy, Adversarial Argumentation, and Embattled Reason. *Informal Logic* **30**(3), 203-234.

Roth, M. (2010). Beyond critical thinking. *The Chronicle Review of the Chronicle of Higher Education*. www.chronicle.com/article/beyond-critical-thinking.

Roth, M. (2021a). Critical feeling and the limits of critical thinking. www.joinexpeditions.com/exps/914.

Roth, M. (2021b). A Focus on critical feeling. *Inside Higher Ed*. www.insidehighered.com/views/2021/03/18/colleges-should-teach-critical-feeling-well-critical-thinking-opinion.

Salmon, M. (2013). *Introduction to Logic and Critical Thinking, 6th ed.* Boston, MA: Wadsworth.

Sanders, K. *et al*. (2017). Folk Pedagogy: Nobody Doesn't Like Active Learning. Proceedings of the 2017 ACM Conference on International Computing Education Research, 145-154.

Scriven, M. (2021). Teaching critical thinking. A radical suggestion. In Blair, J. A. (ed.) *Studies in Critical Thinking, 2nd Ed*. Windsor: Windsor Studies in Argumentation.

Scriven, M. and Paul, R. (1987). Defining critical thinking. *8th Annual International Conference on Critical Thinking and Education Reform*.

Siegel, H. (1988). *Educating Reason: Rationality, Critical Thinking and Education*. New York: Routledge.

Smith, B. O. (1953). The improvement of critical thinking. *Progressive Education* **30**(5), 129-134.

Smith, G. (2002). Are there domain-specific thinking skills? *Journal of Philosophy of Education* **36**(2), 207-226.

Smithson, M. (2016). Human understandings of probability. In Hájek, A and Hitchcock, C. (eds). *The Oxford Handbook of Probability and Philosophy*. Oxford: Oxford University Press.

Stanger-Hall, K. (2012). Multiple-choice exams: An obstacle for higher-level thinking in introductory science classes. *CBE Life Sciences Education* **11**(3), 294-306.

Sternberg, R. (1986). Critical thinking: Its nature, measurement, and improvement. National Institute of Education. http://eric.ed.gov/PDFS/ED272882.pdf

Thayer-Bacon, B. (1998). Transforming and redescribing critical thinking: Constructive thinking. *Studies in Philosophy and Education* **17**, 123-148.

Tibbett, C. (2016). What recent research on diagrams suggests about learning *with* rather than learning *from* visual representations in science. *International Journal of Science Education* **38**(5), 725-746.

Tiruneh, D. *et al*. (2016). Systematic design of a learning environment for domain-specific and domain-general critical thinking skills. *Educational Technology Development and Research* **64**, 481-505.

Turnbull, S. (2017 Nov 22). What's so critical about critical thinking? *Medium*. medium.com/@steveturnbull/whats-so-critical-about-critical-thinking-5a4e4556fd4a

Van Evra, J. (2008). Richard Whately and logical theory. In Gabbay, D. and Woods, J. (eds), *Handbook of the History of Logic*, Vol.4. Amsterdam: Elsevier.

Van Fraassen, B. (1989). *Laws and Symmetry*. Oxford: Oxford University Press.

Van Gelder, T. (2005). Teaching critical thinking: Some lessons from cognitive science. *College Teaching* **53**(1), 41-46.

Verschueren, N. *et al*. (2005). Everyday conditional reasoning: A working memory-dependent tradeoff between counterexample and likelihood use. *Memory & Cognition* **33**(1), 107-119.

Walton, D. (2005). *Abductive Reasoning*. Tuscaloosa: Alabama University Press.

Waller, B. (2001). Classifying and analyzing analogies. *Informal Logic* **21**(3), 199-218.

Whately, R. (1826). *Elements of Logic*. London.

Watts, I. (1726). *Logick: Or the Right Use of Reason in the Enquiry after Truth*. London.

Willingham, D. (2020 Fall). "How can educators teach critical thinking?" *American Educator*, 1-51.

Wojtowicz, Z. and DeDeo, S. (2020). From probability to consilience: How explanatory values implement Bayesian reasoning. *Trends in Cognitive Sciences* **24**(12), 981-993.

Woods, J. *et al*. (2013). *Argument*: *Critical Thinking, Logic, and the Fallacies*, 2nd ed. Toronto: Pearson.

Woods, J. (2021). Abduction and inference to the best explanation. In Blair, J. A. (ed.) *Studies in Critical Thinking, 2nd Ed*. Windsor: Windsor Studies in Argumentation.

Woods, J., Irvine, A., and Walton, D. (2000). *Argument*: *Critical Thinking, Logic, and the Fallacies*. Toronto: Prentice Hall.

Appendix A.

Critical Thinking for Reading Mainstream Science Journalism

In Part 2 I say that Parameter 3 (Comprehension vs Communication) takes a middle value for this CT course. Students learn to construct and articulate reasoning and objections to others′ reasoning. They also learn simply to follow others′ reasoning in order to learn and to decide whether they agree. The content of the course is justified in part by the content of the sorts of serious general journalism that curious people read to become informed and reach judgements about important issues of general concern.

Here I go in some detail through two articles that I think are representative of the sort of writing that should be readable by college students and graduates, and which many students would be interested to read if they felt empowered to do so. The articles, on social media and teen mental health, and the origins of the COVID-19 pandemic, address the sorts of scientific questions of which an informed citizenry in 2022 would have some awareness and understanding. These are serious journalism, broadly classifiable as science journalism, but appear in a mainstream publication (*The Atlantic*) that has a large general readership of non-scientists. Equipping students to read mainstream science journalism like this is partly what warrants Simon Fraser University's designation of this course as a Science and Social Science breadth credit.

Article 1

Jonathan Haidt's article "The Dangerous Experiment on Teen Girls" (2021 Nov 21) in *The Atlantic* presents some of his deep concerns about the effects of smartphones and social media. I quote extensively from the article to show how a full comprehension of it draws on nearly every major topic covered in the course.

The article defends the thesis:

> The available evidence suggests that Facebook's products [in particular Instagram] have *probably* harmed millions of girls.

Haidt announces that he'll be developing an emphatically probabilistic inference. He cannot be certain of his conclusion but uncertainty doesn't mean ignorance. He has collected and analyzed necessarily incomplete and imperfect evidence, which supports his conclusion sufficiently strongly that it make sense to defend it as the thesis of an argumentative essay. Becoming comfortable with uncertainty and probabilistic judgement is a major goal and theme in Units 4 and 6-10.

> If public officials want to make that case, it could go like this:

He announces that he will present an argument (a "case") for the conclusion stated. Beginning with Unit 1, the course focuses on arguments as our most important tool for communicating evidence-based persuasion.

> One major question, though, is how much proof parents, regulators, and legislators need before intervening to protect vulnerable young people.

He reminds us that the probability threshold for action is always an additional consideration and must be based on values, on the costs of acting in different ways, and on the consequences of being right or wrong. This threshold will be debatable in the context of cooperative actions and could be different for individuals acting separately. As I discuss in Sec. 3.2.3, the exercises in Unit 7 incorporate this as a distinct element of the scenarios for students to analyze.

> It is rare to find an 'elbow' in these data sets — a substantial and sustained change occurring within just two or three years. Yet when we look at what happened to American teens in the early 2010s, we see many such turning points, usually sharper for girls… nobody has yet found an alternative explanation for the massive, sudden, gendered, multinational deterioration of teen mental health during the period in question.

He identifies a specific observed fact that calls for explanation. The examples, exercises, and assignments of Units 4 and 7 (general explanations) and 10 (causal explanations) require students to identify the explanans in a story, argument, or piece of news.

> 1 Some have argued that these increases reflect nothing more than Gen Z's increased willingness to disclose their mental-health problems.
>
> 2 Some have suggested that the cause of worsening mental health could be the economic insecurity that followed the 2008 global financial crisis.
>
> 3 Some have suggested that the 9/11 attacks, school shootings, or other news events turned young Americans into "generation disaster."

He seeks and acknowledges competing explanations. Again this is a key part of Units 4, 7, and 10.

> 1 Researchers have found corresponding increases in measurable behaviors.
>
> 2 Why this would hit younger teen girls the hardest is unclear. Besides, the American economy improved steadily in the years after 2011, while teen mental health deteriorated steadily.
>
> 3 Why, then, do similar trends exist among girls in Canada and the U.K.?
>
> Until someone comes up with a more plausible explanation for what has happened to Gen Z girls…

He judges the alternative hypotheses to be unlikely in light of background information. These explanations are difficult to reconcile with other established facts. Identifying, performing, and communicating this procedure is an important part of the exercises in Units 4 and 7. Haidt invokes inference to the best (most plausible) explanation currently available, the subject of the IBE section of Unit 4.

Objections (2) and (3), being phrased as rhetorical questions, also implicitly take the form of Denying the Consequent arguments: If (2) were true, fact F would not be true, but F is true, so (2) is not true. The examples and exercises of Units 3 and 4 have students identify, reconstruct, and compose examples of DC to make an objection, including informal uses that employ a rhetorical question in the conditional premise.

> In 2017, British researchers asked 1,500 teens to rate how each of the major social-media platforms affected them on certain well-being measures.

Haidt doesn't explain that this sample size is the approximate point of diminishing returns for a single survey, where additional data collection is often judged unwarranted because it begins to yield unuseful improvements statistical precision. Although Unit 9 requires no calculations, one major learning objective is for students to grasp in a qualitative way that these sort of sample sizes are usually necessary for high confidence in a precise generalization, and also that they're usually sufficient, that statistics for enormous populations can be reliably estimated from samples of manageable sizes.

> Here's the problem with these studies: Most lump all screen-based activities together (including those that are harmless, such as watching movies or texting with friends), and most lump boys and girls together. Such studies cannot be used to evaluate the more specific hypothesis that Instagram is harmful to girls.

Here the key issue is precision. This is a concept of great and repeated importance in the course — the precision of categories and statements in Units 2-3, of hypotheses and predictions in Units

4-5, and of statistics in Unit 9 (where the fundamental lesson is that a confidence level is always at some degree of precision, with greater confidence coming at the expense of precision).

> It [an overly general analysis] is like trying to prove that Saturn has rings when all you have is a dozen blurry photos of the entire night sky. But as the resolution of the pictures increases, the rings appear.
>
> Andrew Bosworth, a longtime company executive, wrote: ‘While Facebook may not be nicotine I think it is probably like sugar. Sugar is delicious and for most of us there is a special place for it in our lives. But like all things it benefits from moderation.’

These passages use or quote analogies. They each illustrate a point but also function as the premise of argument. The implicit conclusions of these arguments appear to be normative: We ought to use the finer grained analysis. Social media is fine in moderation. Unit 8 is dedicated to the recognition, reconstruction, and evaluation of this ubiquitous type of reasoning.

> Social-media platforms are not like sugar. They don’t just affect the individuals who overindulge.

An important way to probe or criticize analogical reasoning is by considering the strength of the analogy. Haidt objects that the argument’s conclusion is not well supported because the analogy misses an important difference between social media and sugar. The exercises in Unit 8 require students to formulate an objection like this, either their own or one they would anticipate from a reasonable critic.

> Young women were randomly assigned to use Instagram, use Facebook, or play a simple video game for seven minutes. The researchers found that ‘those who used Instagram, but not [those who used] Facebook, showed decreased body satisfaction, decreased positive affect, and increased negative affect.’
>
> Most of the experiments that randomly assign people to reduce or give up social media for a week or more show a mental-health benefit, indicating that social media is a cause, not just a correlate.

He refers to randomized controlled trials, the ideal experiment for demonstrating causation in the social sciences. Units 9 and 10 devote considerable time to explaining the rationale and the procedure. While these units do not require the calculations that a statistics course would, the treatment is not superficial. The exercises use numbers and force students to explain quantitatively what randomization is for and how it can be expected to eliminate confounding factors.

> Children have nevertheless been the subjects of a gigantic national experiment testing the effects of [social media]. Without a proper control group, we can't be *certain* that the experiment [giving Instagram to teenage girls] has been a catastrophic failure, but it *probably* has been.

He refers to the key concept of a control group. Most of Unit 10 is devoted to working through the procedure of a controlled experiment. Students learn the importance of a control group, the risks of informal causal reasoning where inferences are based on assumptions of what would be observed in a suitable comparison.

Article 2

This article is shorter. I include it in this supplement because its language is a better example than is Article 1 of reasoning that may be comprehended more deeply, and more easily, with the benefit of my treatment of Bayes' rule in Unit 7.

Daniel Engber's article "The Lab-Leak Theory Meets Its Perfect Match" (2021 Nov 24) in *The Atlantic* presents the then-close contest between two competing hypotheses about the original human infection by the SARS-CoV-2 virus: direct animal-to-human transmission in the Huanan Seafood Market (sometimes bizarrely called "natural spillover", as though there is something natural about the hellish conditions of live animal storage at the market), and infection from a virus sample, taken from wild bats, being studied at the Wuhan Institute of Virology. The magazine editor supplies the subheadline "Both pandemic-origin arguments depend on coincidence" and much of the article is structured around the semi-rhetorical question "What are the odds?". The reference to unquantified but low probabilities is at the heart of the case Engber presents for each hypothesis, framed as a case against competing hypotheses.

As Engber puts it, summarizing the view of advocates for the laboratory accident hypothesis:

> What are the odds… that a global outbreak of the bat-derived SARS-CoV-2 coronavirus would just happen to begin a few miles from the Wuhan Institute of Virology—the world's largest BSL-4 facility and the research home of the world's leading bat-coronavirus researcher? Now factor in that scientists from this very same facility were involved in efforts—both proposed and carried out—to tinker with the genomes of similar viruses and see whether they got more infectious. And factor in that these same scientists had identified one of the pandemic virus's closest-known relatives in a copper mine about 1,000 miles away.

Here “just happen to” must be read as “if the seafood market or some other competing hypothesis is true”. The probability of an outbreak origin so close to the institute, conditional on the market or other competing hypothesis, is very low. This term, $\boldsymbol{P}(E|\neg H)$, appears in the denominator of Bayes’ rule, and so a low value (significantly lower than $\boldsymbol{P}(E|H)$ in the numerator) makes for strong support for the laboratory accident hypothesis. Combined with a moderate prior probability (based on, for example, awareness of cases of pathogens infecting workers at other ostensibly highly secure labs), this seems to make for a fairly strong case for the laboratory accident hypothesis.

Turning to the market spillover hypothesis:

> What are the odds that, if the virus did emerge from the Wuhan Institute of Virology, the first known cases would show up more than eight miles away, in a market complex across a river from the putative site of any “laboratory-associated incident”? Now factor in that this market complex, in particular, is one of fewer than two dozen in Wuhan—a city of 11 million people—known to have trafficked in live wild mammals as recently as November 2019; and that Chinese health officials had taken an Australian virologist to this same market five years earlier, telling him, as he recalled in a recent interview with Qiu, that it was “a possible place that could trigger future outbreaks”; and also that most of the cases from Huanan were confined to the market’s western section, which, according to the joint pandemic-origins study by China and the World Health Organization, housed seven of the 10 stalls from which wildlife products were sold.

Here the conditional probability $\boldsymbol{P}(E|\neg H)$ is explicit (“if the virus did emerge from the Wuhan Institute of Virology”, this being the only alternative hypothesis and so equivalent to ¬H). The reasoning works the same. Hence the “perfect match” between two competing explanations (although Engber, who probably didn’t write the title, favours the market hypothesis, as does subsequent evidence (Maxmen 2022)).

The COVID-19 pandemic shaped the lives of all the students who’ve taken this course, and who will take it in years to come. Its origin is a fascinating and controversial issue that is potentially hugely important with respect to culpability and prevention. In this article, Engber presents a face-off between the two leading explanatory hypotheses, with the case for each implicitly invoking what I call, in Unit 7, Bayes Lesson #2: “Confirming evidence: More likely if the hypothesis is true, less likely if the hypothesis is false.” These four Bayes lessons encapsulate a non-quantitative, informal version of Bayesianism that equips students to readily comprehend the language and reasoning of a piece of mainstream journalism like this after having worked through and practiced with the quantitative, formal foundation.

Appendix B.

Course Assignments

The course has two short assignments, both generally on what I call scientific reasoning. The first has a writing focus; the second has a more quantitative focus.

Assignment 1

PART 1: Standard Form

Write the three statements included in a disconfirmation argument: Hypothesis (H), Background Assumption (BA), and Evidence (Prediction) (E)

Write an extended deductive argument in standard form with the pattern below.

Main Argument (DS pattern)	Sub-argument (DC pattern)
• (¬H)-OR-(¬BA)	• If H then E unless ¬BA. / If H-AND-BA then E.
• BA *	• ¬E *
¬H	(¬H)-OR-(¬BA)

You are arguing against this hypothesis. You are claiming that H is false.

There are two different ways you could write the conditional premise. Choose one.

To be sure that you have a good H, E, and BA, ask yourself if they make sense together like this:

- H, along with BA, makes you expect E.
- Imagine that H is true but BA is false. Now: it's easy to imagine that E is false.
- E is false. Why? Maybe because H is false. Or maybe because BA is false.

At least one of the premises marked with a star (*) in the patterns above must be information that you have learned from a news article published since 20XX XXX 1.

Part 1 is marked on:

- H, E, and BA make sense together in the conditional premise.
- You have an extended valid deductive argument. The main argument is a DS pattern; the sub-argument is a DC (disconfirmation) pattern.

PART 2: Paragraph

Re-write (the same words) your Part 1 argument in a paragraph or two (max 250 words, not including footnote). Minor wording changes are fine, including saying "probably", "I think", etc. – although these are truth-preserving arguments, if you're not certain that a premise is true, you may not be certain that the conclusion is true.

The paragraph(s) should contain a few additional sentences that introduce and discuss the topic. Some of these statements may support your background assumption, however this does not need to be a deductive argument and you do not need to include it in Part 1. Within your paragraph, **write the statements of your Part 1 argument in bold like this**.

Mark the statement that is supported by a news article with a star at the end like this.* Below your paragraph, include a footnote that gives the news article information in this format: Article Title, News Source, Full Date, Article URL. This is a requirement. Without this reference, your submission is worth no marks.

* "Event Happens", *Gotham Daily News*, 2022 Dec 15, www.gdn.com/2022/12/15/major-event

Write your paragraph word count below the paragraph.

Part 2 is marked on:

- Your paragraph contains the entire Part 1 argument in (almost) the same words.
- Your paragraph(s) is well written.
    - It is a paragraph, not just a list of statements.
    - The extra discussion is relevant and well connected to the deductive argument.
    - It's clear (e.g. from inference indicators) which two statements are conclusions.

Assignment 1: Extended Deductive Argument (Units 3-4)

Sample Assignment

PART 1

Hypothesis (H): Mosquitoes are ecologically important.

Background Assumption (BA): The UN is not concerned only about the economic costs of malaria.

Prediction (E): The UN Convention on Biodiversity does not approve the development of gene drive technology.

Sub-argument

[Premise 1] E if H – unless ¬BA.
The UN Convention on Biodiversity would not have approved the development of gene drive technology if mosquitoes were ecologically important – unless the UN is concerned only about the economic damage caused by malaria.

[Premise 2] ¬E
The UN Convention on Biodiversity approved the development of gene drive technology.

[Intermediate Conclusion] (¬H)-OR-(¬BA)
Either mosquitoes are not ecologically important or the UN is concerned only about the economic damage caused by malaria.

Main Argument

[Premise 3] (¬H)-OR-(¬BA)
Either mosquitoes are not ecologically important or the UN is concerned only about the economic damage caused by malaria.

[Premise 4] BA
The UN is not concerned only about the economic damage caused by malaria.

[Final Conclusion] ¬H
Mosquitoes are not ecologically important.

PART 2

Although environmental activists are very concerned about a new genetic technology being designed to kill off the mosquitoes that spread malaria, we can be confident that **mosquitoes are not ecologically important**. The new technology, called a "gene drive", involves genetically altering mosquitoes to eliminate females (the ones that bite) or to make females infertile. The activists fear that killing off mosquitos will disrupt the ecosystem. But **the United Nations Convention on Biodiversity** looked carefully at the risks and benefits of gene drives and **decided to approve their development***. **They would not have approved it if mosquitoes were ecologically important – unless, at any rate, the UN is concerned only about the economic damage caused by malaria**. So **either mosquitos are not ecologically important or the UN is concerned only about the economics**. But while the UN certainly is concerned about the terrible economic toll of malaria in societies where people are constantly sickened by it, the Biodiversity Convention represents a serious commitment to preserving plant and animal species and to mitigating risks of ecological disasters such as the collapse of a food chain. **It's very unlikely that the UN is concerned only about the economic costs of malaria.** The activists are well-intentioned but don't give the UN enough credit.

[Word count: 208 words]

* "The Sci-fi Technology Tackling Malarial Mosquitos", *BBC News*, 2022 May 23, www.bbc.com/news/business-61505102

Assignment 2

PART 1: Observation, Question, and Hypotheses

a) Observation (25 words max not including citation)
State a fact or event reported in a news article published since 20XX XXX 1.
Cite this news article (Title, Date, Publisher, URL) with a note*.
* "Interesting Event Happens", 2022 Dec 15, *The Gotham Herald*
(www.gothamherald.com/2022/09/15/world/interesting-event-happens.html)

b) Discussion (100 words max)
Provide some relevant details about the observation. You may use one additional source.

c) Explanatory Question (25 words max)
Ask a specific question about your observation. This could be simply “Why did [event] happen?”.

d) Hypothesis 1 and Hypothesis 2 (2 × 25 words max)
State two alternative answers (H1 and H2) to your explanatory question.

- Here *competing* means common sense tells you it’s highly unlikely that H1 and H2 are both true.

e) Prior Probabilities (200 words max)
Estimate the probabilities ***P***(H1) and ***P***(H2).

- These numbers are your own judgements of how likely H1 and H2 are, based on sections (a) and (b) and on any other factors you think are relevant (e.g. simplicity, an competing hypothesis H3, etc.).

Say why the probabilities you give are reasonable.

- You do not need to defend the precise numbers you give, only why you think each probability is generally correct, e.g. explain why it is 80% rather than 50%, not 80% rather than 75%.

PART 2: New Evidence

a) New Evidence for H1 (25 words max)
State some possible new evidence E that would either confirm or disconfirm H1.

- *Possible* means that it could happen/exist/be true (E) or not happen/exist/be true (¬E).
- *New* means that ***P***(H1) was not based on E.
- It’s fine if it would be difficult to discover E or ¬E as long as it could be discovered.

b) Evidence Strength (200 words max)
Estimate the probabilities ***P***(E|H1) and ***P***(E|¬H1). For both E and ¬E, calculate the evidence strength and describe it as strongly/moderately/weakly confirming/disconfirming.
Say why the evidence strength you give for E is reasonable.

PART 3: Updating H1 Probability

a) Bayes Box
Draw a Bayes Box for H1 and E.
Write your visual estimates of the updated odds of H1 given E, and of the updated odds of H1 given ¬E.

b) Bayes' Rule

Imagine you learn that E. Write Bayes' Rule for ***P***(H1**|**E) and calculate the updated probability. Imagine you learn that ¬E. Write Bayes' Rule for ***P***(H1**|**¬E) and calculate the updated probability.

- Check that your calculations match your estimates from the diagram.

c) Hypothesis 2 (75 words max)

Imagine you learn that E and you update the probability of H1. Estimate and briefly explain the new probability of H2.

- **Your estimate is based on the "available" probability for H2 and a roughly estimated (not calculated) evidence strength of E for H2.**

Appendix C.

Course Manual

Description:

This document functions as the course textbook, distributed to students as a free PDF. Unit numbers throughout this project refer to the units of this document.

Filename: Critical_Thinking_Course_Manual.pdf

Appendix D.

Thinking about Categories and Counterexamples

Description:

This video demonstrates and discusses two excerpts from the class slides for Unit 2 Category Logic. The first excerpt is from the initial discussion of the concept of a category. The second excerpt is from a sequence at the end of the unit that models a disagreement over a counterexample.

Filename: Categories_Counterexamples.mp4

Appendix E.

Teaching for Deep Structure

Description:

This video demonstrates and discusses two excerpts from the class slides for Unit 9 Generalizing from a Sample. The first excerpt shows an example of how it can be very difficult for students to see the deep structure of a problem even with numerous clues about how to see it. The second excerpt shows an attempt to teach for deep structure (the concepts of precision and confidence) with an example very different from those that came before it.

Filename: Samples_Precision_Confidence.mp4

Appendix F.

Constructing Bayes' Rule for Scientific Reasoning

Description:

This video shows a Zoom class taught in an on-line section of the course. This sequence of slides introduces Bayes' rule. (Detailed examples follow.) The long runtime (92 minutes) reflects a relaxed and conversational class (in a course scheduled four hours per week) devoted entirely to making sure as many students as possible feel comfortable with what we're doing. In a course scheduled in the standard way at Simon Fraser University (100 minutes lecture plus 50 minutes tutorial per week), I would cover this same material in half the time (i.e. a single 50 minute class).

Filename: Bayes_Rule_Zoom_Class.mp4