CRITICAL THINKING:
CHALLENGES, POSSIBILITIES, AND PURPOSE

by:

Marvin S. Cohen and Eduardo Salas

Cognitive Technologies, Inc.
4200 Lorcom Lane
Arlington, VA 22207
(703) 524-4331

and

Sharon L. Riedel

U.S. Army Research Institute
Field Unit-Leavenworth
Ft. Leavenworth, KS 66027

March 2002

Contract No. DASW01-00-C-3010

TECHNICAL REPORT 02-1

CRITICAL THINKING: CHALLENGES, POSSIBILITIES, AND PURPOSE

by:

Marvin S. Cohen and Eduardo Salas

Cognitive Technologies, Inc.
4200 Lorcom Lane
Arlington, VA 22207
(703) 524-4331

and

Sharon L. Riedel

U.S. Army Research Institute
Field Unit-Leavenworth
Ft. Leavenworth, KS 66027

March 2002

Contract No. DASW01-00-C-3010

REPORT DOCUMENTATION PAGE

FORM APPROVED OMB NO. 0704-0188

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	March 2002	Final, January 2000-December 2000

4. TITLE AND SUBTITLE
Critical Thinking: Challenges, Possibilities, and Purpose

5. FUNDING NUMBERS
Contract No.:

DASW01-00-C-3010

6. AUTHOR(S)
Marvin S. Cohen, Eduardo Salas, and Sharon L. Riedel

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Cognitive Technologies, Inc.
4200 Lorcom Lane
Arlington, VA 22207

8. PERFORMING ORGANIZATION REPORT NUMBER

Technical Report 02-1

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Research Institute
Fort Leavenworth, KS 66027-0347

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

Dr. Sharon L. Riedel was the Contracting Officer's Representative.

12A. DISTRIBUTION/AVAILABILITY STATEMENT
Distribution authorized to U.S. Government Agencies only; report contains proprietary data produced under SBIR Contract No. DASW01-00-C-3010. Other requests shall be referred to the performing organization in Block 7 of this form.

12B. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words):

There is widespread interest in critical thinking in the Army and elsewhere, as a set of skills for handling complex, novel, and information-intensive tasks for which initiative is required. Our objectives were to (i) develop a general framework for understanding critical thinking by individuals and teams, and (ii) outline a new, integrative theory of critical thinking based on that understanding. We contrast two competing paradigms. Critical thinking has traditionally been conceptualized from an *internalist* point of view, which locates its validity in rules meant to fit the contents of an individual consciousness. From the *externalist* point of view, critical thinking estimates the reliability in real environments of cognitive processes that produce intellectual products. Our theory integrates elements of internalist and externalist paradigms in a coherent way. Critical thinking skill requires coordination of three different perspectives: proponent, opponent, and judge. The theory synthesizes research in three areas: (1) Cognitive theories according to which alternative *possibilities* are represented by mental models. (2) Normative models of critical discussion in which a proponent defends a claim against *challenge* by a critic. (3) Assessments by a judge about the reliability of cognitive processes or dialogue strategies for achieving real-world *purposes* in a timely way.

14. SUBJECT TERMS
critical thinking, decision making, situation awareness, mental models, dialogue theory, informal logic, metacognition, teams

15. NUMBER OF PAGES
291

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	None

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI std. Z39-18
298-102

ACKNOWLEDGMENTS

We were pleased and fortunate to receive the assistance of a panel of ten distinguished consultants in fields relevant to critical thinking. The panel served as a source of valuable and stimulating advice on a variety of issues discussed in the report (but is in no way responsible for its limitations). The following individuals participated: Jonathan M. Baron (Department of Psychology, University of Pennsylvania), Alan M. Dershowitz (Harvard Law School), Robert H. Ennis (Philosophy of Education Emeritus, University of Illinois at Urbana-Champaign), K. Anders Ericsson (Department of Psychology, Florida State University), Diane F. Halpern (Department of Psychology, California State University), Philip N. Johnson-Laird (Department of Psychology, Princeton University), Karen Kitchener (College of Education, University of Denver), Barbara M. Koslowski (Department of Human Development, Cornell University), Lokendra Shastri (International Computer Science Institute, Berkeley), and Robert J. Sternberg (Department of Psychology, Yale University). We also thank Jared Freeman (of the Aptima Corporation) for his help.

CRITICAL THINKING: CHALLENGES, POSSIBILITIES, AND PURPOSE

EXECUTIVE SUMMARY

Research Requirement:

There is widespread interest in critical thinking in the Army and elsewhere, as a set of skills for handling complex, novel, and information-intensive tasks, especially in situations that demand initiative and independent thought. Questions arise, nonetheless, about the potential usefulness of training critical thinking skills for use on the battlefield: Will it take too much time, undermine the will to fight, supplant experience, stifle innovation, or disrupt coordination? Unfortunately, the current state of the field of critical thinking does not provide ready answers to these questions. Current critical thinking textbooks tend to include an eclectic mix of ideas and methods that borrow from formal and informal logic, probability theory, decision theory, cognitive psychology, communication theory, rhetoric, and others. The various textbooks and approaches do not provide a framework that integrates these competing approaches in a theoretically adequate or practically useful way. Moreover, there is very little empirical research on critical thinking in time-sensitive domains such as battlefield tactical decision making. A better understanding of critical thinking is needed so that the Army can make well-founded choices regarding the design of training and instruction, identify additional research needs and opportunities, and realize the potential benefits of enhanced battlefield critical thinking skills.

Procedure:

The objectives of the research were (i) to develop a general analytical framework for understanding critical thinking and evaluating alternative approaches, and (ii) to outline a new, integrative theory of critical thinking based on that understanding. These objectives are reflected in Parts I and II of the report, respectively.

In pursuit of the first objective, we reviewed the literature in critical thinking and in fields from which it draws such as informal logic, epistemology, logic, decision making, and cognitive psychology. In Part I we addressed a series of issues:

- What claims are made for the utility of critical thinking? What obstacles stand in the way of realizing that utility? (Chapter 1)
- What does it mean to define critical thinking? What types of definition are possible? (Chapter 3)
- How has critical thinking in fact been defined? What are the shared and non-shared features of current definitions? (Chapter 4, Appendix A)
- What are the major differences in underlying assumptions in approaches to critical thinking? What implications do these differences have for the shape of a critical thinking theory? (Chapter 5)
- What specific critical thinking paradigms have been proposed? How do they vary? (Chapter 6) What are the detailed strengths and weaknesses of informal logic as a component of critical thinking? (Appendix B)

The framework that emerged from these questions guided our work on the second objective, the development of an integrated theory of critical thinking. In Part II, we do the following:

- Lay out a theory of critical thinking (Chapter 7),
- Make a case for the new theory by analyzing its relationship to traditional and contemporary theories of knowledge and reasoning (Chapters 8, 9, and 10; Appendix B)
- Apply the new theory to the problem of training and assessing critical thinking skills in teams (Chapter 11)
- Evaluate the usefulness of critical thinking training in the Army battlefield domain in light of the new theory (Chapter 12)

Findings:

In Part I, we reach the following conclusions:

- It is often claimed that critical thinking skills have grown in importance as a result of increased problem complexity, decentralization of organizational structure, and more frequent high stakes decisions. In the Army battlefield context, however, doubts about its usefulness arise due to potential demands on time and training resources, and the possibility that it will stifle innovation or dilute the effects of leadership and experience. (Chapter 1)
- There are three complementary levels at which critical thinking can be studied and defined: normative, cognitive, and applied. The cognitive level can be divided into processes, mechanisms, and their interaction via cognitive faculties. Each of these levels affects the others in important ways (Chapter 3)
- Definitions of critical thinking in the literature vary in part because of their varying emphasis on normative, cognitive process, cognitive mechanism, and applied levels. A common core of current definitions might be that critical thinking is the deliberate evaluation of intellectual products in terms of a standard. Definitions vary with respect to the products to be evaluated, the standards to be used, and the processes and mechanisms that carry the evaluation out. (Chapter 4)
- These differences can largely be accounted for in terms of the competition between two high-level paradigms. Critical thinking has traditionally been conceptualized from an *internalist* point of view, which sees it as taking place within the consciousness of an individual. Rational justification consists in the evaluation of a static set of beliefs through the application of universal (e.g., logical) standards. Cognitive processes and strategies are unimportant since only the information present in the mind at one time is relevant. From the *externalist* point of view, by contrast, evaluation is a matter of estimating the reliability in a real environment of the cognitive processes that produced an intellectual product. Externalist evaluation is highly context-dependent, the relevant processes may be domain-specific, and intellectual products other than beliefs may also be critically evaluated. Cognitive processes that identify biases and fallacies, expose views to challenge, and actively seek information may increase overall reliability in

particular circumstances., But critical thinking is not necessary for rationality: In some circumstances, intuitive or recognitional processes may be more reliable. From the externalist point of view, critical thinking skill includes not only cognitive processes, but also enduring traits or dispositions to adaptively select strategies that have proven reliable. (Chapter 5)

- Mid-level paradigms for critical thinking include approaches like operations research, decision theory, formal logic, informal logic, dialogue theory, bounded rationality, naturalistic decision making, and rhetoric. Differences among these can be understood along two dimensions: whether they admit the relevance of how people actually make decisions to judgments of how they ought to make decisions, and whether they adopt an externalist or internalist stance toward the grounds for an evaluative judgment. (Chapter 6)

In Part II, we reach these conclusions:

- We propose a theory of critical thinking that integrates elements of internalist and externalist paradigms in a consistent way. Critical thinking skill requires coordination of three different perspectives: proponent, opponent, and judge. To understand these three different roles, the theory draws on and synthesizes research in three areas: (1) Cognitive theories according to which alternative *possibilities* are represented by mental models and reasoning is accomplished by manipulating mental models. (2) Normative models of critical discussion in which a proponent must defend a claim against *challenge* by an opponent or critic. (3) Assessments by a judge about the reliability of cognitive processes for achieving external *purposes*. Dialogue theory provides a bridge between internal and external points of view, since critical thinking dialogues take place within an individual or among different individuals. (Chapter 7)

- Standard approaches to critical thinking are heavily influenced by classical and contemporary foundationalism, the view that knowledge is built up cumulatively one step at a time from solid foundations. From the point of view of our theory, this approach places constraints on critical dialogue that are not always appropriate. Traditional views unduly constrain critical thinking dialogue. (Chapter 8) A detailed examination of informal logic provides support for this conclusion. (Appendix B)

- Mental models, or stories, as well as network models of underlying knowledge, are central to a more realistic understanding of critical thinking. Stories and mental models are evaluated in part in terms of coherence. Ultimately coherence can be analyzed in terms of the number and nature of the questions a story answers. Coherent models must be built and maintained by highly flexible question and answer strategies in critical dialogue. (Chapter 9)

- Ultimately, the value of a critical thinking strategy is determined by its success in achieving real-world goals under the relevant conditions. Instead of viewing the process “from the inside” (e.g., what reasons do I have for this conclusion? Can I answer this objection?), the external point of view looks more generally at the record of success of this type of strategy in similar circumstances in the past. Both points of view are necessary, and they complement one another. The external

point of view determines what cognitive strategy or dialogue type is appropriate and when and how it should be terminated and a decision reached. Problems with the externalist framework can be handled by acknowledging that it reflects a task-relative point of view. (Chapter 10)

- Team decision making depends on shared mental models of the task, the situation, and the communicative processes within the team that create and maintain such shared knowledge. A key practical application of the critical thinking theory, therefore, is to team decision making. Rules for the conduct of each stage of critical discussion, taken together, provide a normative model for team problem solving. The theory can be used to develop training objectives, training content, and assessment measures. (Chapter 11)
- The critical thinking theory provides preliminary answers to challenges raised in Chapter 1. The theory provides two crucial types of flexibility: (i) There is an array of dialogue types that differ in the intensity with which underlying assumptions are probed and which are suited to different contexts. (ii) The judge, adopting an external point of view, determines what strategy will most reliably achieve the real-world objective, including among the options non-deliberative processes such as recognition-based decision making.

Utilization of Findings:

An adequate theory of critical thinking, with both theoretical and applied dimensions, is a key condition of progress in the development of critical thinking training and support. Such a theory is needed to guide the application of critical thinking principles to Army battlefield contexts as well as to a variety of other domains.

The new theory of critical thinking combines theoretical soundness with practical utility. At the practical level, it lends itself directly to operationalization: concrete specification of the practices that make up successful critical thinking in different contexts. These specifications in turn serve as the objectives of critical thinking training or decision support. Each of the three components brings with it criteria for success and methods for the identification of errors. The theory should help us specify critical thinking objectives, develop training material, and measure success. The ultimate result should be better decision making by both individuals and teams.

TABLE OF CONTENTS

TABLE OF FIGURES

PART I: THE PROBLEM

1. THE PURPOSE OF CRITICAL THINKING

Why Study Critical Thinking?

The present research arises out of the expectation that improved understanding of critical thinking skills and training to improve those skills will result in better decision making by Army battlefield command teams. In critical thinking a cognizer's beliefs, plans, inventions, practices, or other creations are challenged, defended, replaced, and/or modified in order to achieve some objective. We think critically when we ask ourselves or others *questions*, such as: How do I know this? Why did I decide to do this? Are the reasons for this belief adequate? Will this action or design achieve its intended effect? What situations can I imagine in which this belief is false, or this plan fails? Is there something else I need to know or think about? What arguments are there against this conclusion or course of action? Is there a better hypothesis or plan? What is my real purpose, and am I addressing the issues that really concern me?

This report describes progress on two objectives: First, to draw a map that displays key features of the critical thinking terrain, including the positions of competing points of view on what critical thinking is, how they view one another, their concealed assumptions, obstacles to progress, and promising avenues of approach. Second, to use that map to advance toward a new theory which counters weaknesses in other approaches while systematically integrating their insights. The resulting theory is a perspective on critical thinking as a dialogue that explores alternative possibilities by asking and answering questions, and which takes place under the constraints of context-specific goals. To maximize relevance to Army concerns, our focus will be on the use of such dialogue skills for the achievement of *practical* objectives in a *timely* way.

Both the map and the theory should:

(i) Help the Army and others make well-founded choices in the design of training and instruction. Such training and instruction should improve performance on the battlefield and elsewhere through better critical thinking.

(ii) Help the Army and others identify additional research needs and opportunities that promise significant payoffs in critical thinking training and, ultimately, in real-world outcomes.

(iii) Provide a general way of organizing a field that is both relatively new and multi-disciplinary. The framework and theory provide a scaffolding within which both theoretical ideas and applied proposals about critical thinking can be compared and better understood. Hopefully, it will prove useful as a stimulus for other researchers.

The critical thinking field is currently fragmented at both the strategic and the tactical levels. Critical thinking involves the willingness and ability to *question* unreflective beliefs and accepted practices. But theories diverge about the overall strategy that such questioning serves. Critical thinking started out as a weapon against superstition and dogma, but questions and answers were also expected to lead step by step to certainty about the real world. In modern times, on the other hand, the bedrock under these steps of reasoning seems less solid, and the goal of absolute certainty has been discredited even in physical science and logic. As a result, some of today's theorists have moved to the other extreme, concluding that truth is internal to a belief system and that every such system rests on arbitrary assumptions. The role of critical thinking is to question but *not* to provide firm answers, to dispel certainty but *not* to restore

justified confidence. The implicit assumption is that no position is better or worse than any other; each can be judged only with respect to its own, internal standards. That conclusion is an unacceptable strategic basis for applying critical thinking to real tasks in realistic contexts.

The answer to the strategic problem, we suggest, depends on seeing critical thinking from both an *inside* and an *outside* perspective. The outside perspective lies in the treatment of critical thinking strategies as adaptations to real environments, selected and shaped by their likelihood of producing successful outcomes under prevailing conditions. The inside perspective lies in the decision-guiding rules that critical thinkers follow in order to envision and evaluate alternative possibilities and to construct internally coherent mental models of the situation. From the internal point of view, critical thinking is a rule-governed question-and-answer dialogue about alternative possibilities. But critical thinking is an internal process carried out to achieve an external result. Critical thinking by its very nature demands a balance between skepticism and confidence.

On the tactical level, current critical thinking textbooks tend to include an eclectic mix of ideas and methods that borrow from formal and informal logic, rhetoric, probability theory, decision theory, cognitive psychology, communication theory, and others. In some obvious sense, all of these serve as tools of critical thinking. But the various textbooks and approaches do not provide a framework that integrates these competing approaches at a meaningful level of detail. The theory we propose, on the other hand, brings together three top-level elements:

- *challenge* through question and answer dialogue,
- alternative *possibilities* of various types whose exploration and evaluation is the proximal goal of that dialogue, and
- a *purpose* whose reliable achievement is the ultimate measure of success of the dialogue.

At the tactical level, therefore, the theory draws on *and integrates* three broad research areas. The first is work in cognitive psychology on reasoning, decision making, and problem-solving. The central lesson we extract from this work is that significant errors in logic, probability, problem-solving, or creative thinking occur when cognizers do not adequately explore relevant alternative possibilities (e.g., Hastie & Dawes, 2001; Dawes, 2001). These alternative possibilities can be represented as *mental models* (Johnson-Laird & Byrne, 1991). The theory proposes *dialogue* as a framework for understanding the strategies people use for constructing and evaluating such models. In a wide range of tasks, alternative mental models are explored by asking and answering questions within a specifiable framework of rules and expectations, i.e., by exercising skills of *dialogue*. Thus, the second key area is work on dialogue theory by informal logicians (e.g., van Eemeren & Grootendorst, 1992, and Walton, 1998). A dialogue is defined as a verbal interaction of a specific type conducted by parties who play different roles and governed by rules that are appropriate for achieving the proximate objective of that type of dialogue, e.g., persuading another person (or oneself) of a point of view, choosing a course of action, or negotiating a resolution of competing interests (Walton, 1998). Errors or lack of skill in the execution of a dialogue causes important alternative possibilities to be overlooked. Reasoning is a dialogue with oneself, and conversely, a dialogue is a collaborative process of reasoning.

Since there are alternative types of dialogue to choose from, since the process of challenging mental models and generating alternative possibilities could in principle go on forever, and since critical thinking is sometimes not pragmatically appropriate at all, critical

dialogue must be placed in a larger, external context. A dialogue that achieves its proximal objective (e.g., resolving a difference of views, selecting an action, or reaching an acceptable balance among competing goals) might nevertheless fail to contribute to the ultimate purpose of the task or activity within which it is embedded. Decisions about whether to conduct a critical thinking dialogue, what kind of dialogue to conduct, whether the rules have been violated, and when to bring the dialogue to an end are based on the *reliability* of different dialogue types or strategies for achieving the real-world objectives of the participants under the current conditions and within the time available. The third key area of research, then, is empirical and theoretical work on adaptive cognitive strategies that are reliably associated with expertise or successful performance (e.g., Ericsson & Smith, 1991; Payne, Bettman, & Johnson, 1993). The three key concepts in the theory – dialogue, mental models, and reliability – thus provide the foundation for a deep and broad theoretical synthesis.

This view of critical thinking, as reliably effective dialogue about alternative possibilities, aims to combine theoretical soundness with practical utility. At a practical level, it lends itself directly to operationalization, i.e., concrete specification of the practices that make up successful critical thinking in different contexts. These specifications in turn serve as the objectives of critical thinking training or decision support. Each of the three components brings with it a set of inter-related criteria for success and a method for the identification of errors. The theory should help us specify critical thinking objectives, develop training material, and measure success.

Critical thinking is not a substitute for knowledge and experience in a particular field, such as medicine, law, business, or military tactics. For one thing, such knowledge is required in order to generate and evaluate alternative possibilities. But critical thinking can be a powerful knowledge-amplifier and a crucial element of learning in any field. A critical thinking dialogue challenges habits and settled beliefs, exposes hidden assumptions, helps identify and fill gaps in knowledge, brings out alternative approaches that might never have been considered, speeds up learning, and keeps us on track toward achieving our goals. Critical thinking skill is creative: It requires the ability to know when to follow a gut feeling and when to put it on hold, when to fashion new solutions and when to adapt old ones, and in general to use our knowledge effectively while enlarging it. It is the ultimate goal of a critical thinking theory to illuminate both how these skills operate and how they may be improved.

When Is Critical Thinking Useful?

Will critical thinking be useful in Army decision making? The answer will come via experimental tests toward the end of this research rather than at its beginning. It makes sense, however, to start by looking at what proponents of critical thinking have *claimed* about its usefulness. It turns out that a small set of themes appears over and over in the prefaces and introductions of the dozens of critical thinking textbooks in print. Critical thinking is becoming more important because of:

A. Growing problem difficulty

1. Increasing complexity of problems
2. Changing nature of problems
3. Information overload

B. Decentralization of social and organizational structure

4. Increasing responsibility, hence the need for initiative
5. Increasing need to participate in teams with diverse membership
6. Increasing need for independent thinking

C. More high stakes decisions

7. Increasingly important public policy issues
8. Personal decisions in an increasingly competitive career environment

Many authors cite trends in some or all of these categories.[1]

---

[1] Here are some briefs for critical thinking which give a flavor of textbook rhetoric. Numbering is ours and corresponds to the list in text:

> …in a world of (2) accelerating change and (1) complexity, a new form of thinking and learning is required …. The economic well-being of the future will require the (4) intellectual empowerment and freedom of ordinary, not just extraordinary people. (Paul, 1993, p. v)

> We live in what has been called the (3) Information Age because of the many messages that we receive daily from newspapers, magazines, radio, television, books, and the Internet….in a (6) democratic society…we need to know how to understand and evaluate the information that comes our way. (Diestler, 1998)

> We are on the receiving end of (3) an enormous amount of argumentation demanding our acceptance or support…. These things are particularly important in a (6) democracy, the success of which requires that all of us make (7) significant decisions about social and economic issues that determine public policy. (Groarke, Tindale, & Fisher, 1996, pp. xiii-xiv)

> …At our workplaces we seek to exercise (6) democratic control over workplace functions and organization and to (4) take initiative in charting (2) new directions and in designing the form and content of our activities…Politically, we value (4) freedom, we practice (6) democracy, we encourage a tolerance of (5) diversity… (Brookfield, 1987, p. ix)

> Each of us is bombarded with (2) information… In all areas of knowledge there are issues about which (1) experts in those fields disagree… (Browne & Keeley, 1994)

> For the first time in the history of the human race, we have the (7) ability to destroy all life on earth. The decisions that we make as individuals and as a society regarding the (7) economy, conservation of natural resources, and the development of nuclear weapons will affect future generations of all people around the world. We are also called upon to make decisions on a wide range of (8) important local and personal topics (Halpern, 1996: p. 2).

> Competency in critical thinking is a prerequisite to participating effectively in (7) human affairs, (8) pursuing higher education, and (8) succeeding in the highly competitive world of business and the professions. (Freeley & Steinberg , 2000: p. 1)

> Men would sooner (8) die than think. In fact they do. (Bertrand Russell)

The following quote gives an unusually complete rationale for critical thinking training, perhaps because it is from the point of view of critical thinking in a specific domain, the nursing profession:

> Have you noticed that nothing seems simple anymore – that as we improve and progress, life only seems to get more (1) complicated?…People live longer with more chronic and complex problems providing (2)

Do These Conditions Apply in the Army?

All of these trends seem operative in the Army. Perhaps not coincidentally, there is a growing interest in critical thinking among Army instructors and researchers, as well as doctrine developers and planners. This interest is warranted both by (A) problem difficulty, exemplified by the complexity and changing character of military planning and operations; (B) decentralization of the organizational structure, exemplified by the demands of leadership, coordination, and initiative within every echelon; and (C) high stakes personally, organizationally, and for the nation as a whole. The direction of change in the Army promises to exacerbate all the relevant factors. These changes include the growing complexity of military tasks, the rapid evolution of technology and missions, the flood of information unleashed by the new technology, increasing diversity of membership in national and allied forces, and the essential role of tactics that rely on initiative by local teams.

Clearly, a good case can be made that critical thinking is an important Army battlefield skill, and that its importance is likely to increase. But it is important to get beyond rhetoric and surface compatibility between critical thinking and Army needs – to evaluate the prospects in more detail. Unfortunately, most textbooks stay at a general level: After discussing the broad trends listed above, they go directly to the presentation of critical thinking techniques (e.g., formal logic, informal logic, and probability) without explaining the connection between the techniques and the challenges they allegedly address. But we need to know *how* critical thinking will help address the various challenges and under what specific conditions. For example, it is not obvious that logic or decision theory will be of much use in highly complex and/or information-intensive problems.

Let us play the devil's advocate. The current state of critical thinking research and instruction leaves some important questions unanswered:

<u>Is critical thinking consistent with tactical battlefield constraints?</u>

- Will critical thinking on the battlefield take too much time? Would that time be put to better use gaining a jump on the enemy?
- Will critical thinking result in a loss of the confidence necessary for decisive leadership and action? Will it undermine the "will to fight"?

<u>Is critical thinking consistent with other battlefield skills?</u>

- Will critical thinking skills trump experience or leadership qualities on the battlefield, which might in fact lead to better decisions?
- Will critical thinking be too "critical"? Will it stifle innovation or the development of new tactics and techniques?

---

new challenges. Computers give us instant access to (3) vast knowledge stores, making it hard to find what it is *we* need. In communities, schools, and especially in the workplace, we're all expected to accept (4) more responsibilities, work with (5) diverse teams, and make (6) more independent judgments and decisions. (Alfaro-LeFebre, 1999; p. 4)

Is critical thinking appropriate for military organizational structure?

- Will critical thinking encourage inappropriate initiative? Will it disrupt the chain of command and degrade coordination and synchronization on the battlefield? Put another way, is the Army too centralized and hierarchical for critical thinking to flourish?
- Will critical thinking hinder the development of trust among members of diverse, multi-cultural teams because it is "Western, masculine, individualistic, adversarial, and coldly rational" (Atkinson, 1997; cited in Davidson, 1998).

Will critical thinking fit into Army training?

- Are there “right answers” in critical thinking? If so, isn’t this just a new phrase for teaching doctrine and tactics, which we already do? If not, what good are skills that can’t be evaluated? How can we know they will improve performance?
- Will critical thinking instruction consume too much training time? How will we persuade instructors to provide that time? Does critical thinking require technical training in logic or decision theory? Does it require stand-alone courses? How will we persuade students to devote their time to the study of critical thinking?

One of the purposes of this report is to provide a basis for answering these questions. If the result of this evaluation is to be positive, it requires clarification of what critical thinking means both in general and in specific Army settings.

2. AN EXERCISE IN EVALUATION

Do I Know It When I See It?

To train critical thinking, we must be able to recognize it when it occurs and to evaluate its quality. The main purposes of this chapter are to (i) introduce an example, and (ii) help readers elicit their own intuitions about critical thinking before reading further. It presents a brief recognition and evaluation exercise, in the form of a short dialogue between two Army officers. After each segment of dialogue is presented, the reader is invited to consider if critical thinking took place, and what (if anything) was good about that segment of dialogue and what was bad about it from the point of view of critical thinking skill. The exercise will provide a simple concrete example, which we will refer to in later discussions. After each segment, we will suggest some possible answers. They are not meant to be definitive, but only to raise some issues that will reappear in subsequent discussion.

Segment 1

MAJ South: *If the enemy attacks, do you think they'll come through the northern pass or the southern one?*

MAJ North: *It won't be the southern one, that's for sure.*

Sud: *Why not?*

Nord: *Because they haven't got any bridging equipment.*

Sud: *So, you think it's the north then?*

Nord: *Yep.*

Now, before continuing, does critical thinking occur in this dialog? If so, what are its good points and what are its bad points? Can you judge whether it is good or bad critical thinking?

****

Here are some possible good points, from the point of view of critical thinking:

- Sud **recognizes her own uncertainty** and attempts to handle it, by **soliciting another person's opinion**.
- After hearing Nord's opinion, Sud **reserves judgment** and **asks for a reason** for Nord's conclusion.
- Nord **defends her own position** by **identifying relevant evidence** (lack of bridging equipment) and the **direction of support** provided by the evidence (against attack in the south).
- Sud asks Nord to **clarify** the conclusion of Nord's argument, and either to **commit** explicitly to it, or **express any relevant qualifications or doubts**.

But here are some possible bad points:

- What did Sud **presuppose in the initial question**? Hint: What reason does Sud have to think he has considered **all the relevant hypotheses** about direction of attack? (Might the enemy attack through both passes? Might it bypass both passes, e.g., by using air assault?)
- Has Nord **jumped to a conclusion** prematurely? Does the evidence he provides appear to be **sufficient** to rule out an attack through the south?**Segment 2**
- MAJ Sud: *Well, I don't agree. They don't have any artillery in the North, and they would never attack without it.*

Before continuing, does critical thinking occur in this brief segment? If so, what are its good points and what are its bad points? Can you judge whether it is good or bad critical thinking?

****

Here are some possible good points:

- Sud **challenges** Nord's conclusion rather than accepting it without question.
- Sud **identifies evidence that conflicts** with Nord's conclusion (the location of artillery).
- Sud gives Nord an **opportunity to defend** Nord's conclusion against the challenge .

Here are some possible bad points:

- Has Sud **jumped to a conclusion**? Sud has given **no reason** to support the claim that the enemy would not attack without artillery. Is it proper for Sud to rely on **common knowledge** that the enemy will not attack without artillery, rather than **defending** this claim?
- Sud has **not responded directly to Nord's argument** regarding bridging equipment. Sud has simply gone on to present an argument for the position Sud favors. Is this OK?
- Is it appropriate for Sud to use **absolute words** like "never" in this context? Is this a way of **suppressing uncertainty**, or **discouraging counterarguments** from others?

Segment 3

MAJ Nord: *But don't we have reports that the enemy is developing longer-range artillery?*

MAJ Sud: *True, but I don't recall any indications that they've deployed the new systems yet.*Before continuing, does critical thinking occur in this segment? If so, what are its good points and what are its bad points? Can you judge whether it is good or bad critical thinking?

****

On the positive side:

- Nord counters Sud's argument directly by **challenging the inference** from Sud's evidence (artillery location) to Sud's conclusion (location of attack).
- Nord **exposes an implicit assumption** about artillery range in Sud's inference.

- Sud responds directly to Nord's challenge by providing **evidence in defense of the assumption**.
- Sud defends her own assumption by **exposing an assumption** in Nord's challenge (that the enemy has deployed the longer-range artillery).

On the negative side:

- Nord does not spell out Nord's full argument to Sud. For example, it **relies on an implicit assumption** like the following: *If the new artillery has a range longer than 30 km., they wouldn't have to move it in order to use it in the north.* Is it OK not to state one's full argument? (Nord's earlier argument, based on bridging equipment, also involved implicit assumptions, e.g., that *There are unfordable rivers in the south but not in the north.*)
- In raising the possibility of longer range artillery, do you suspect that Nord is unreasonably **discounting evidence that conflicts with his original view**?
- Is Sud also **explaining away evidence that conflicts** with his view?
- If you said yes to the latter two questions, how do you reconcile your answers with the positive points above? Can you challenge another's evidence without appearing to unduly favor your own position?

Can Context Be Ignored?

The reader may rightfully complain that this is an extremely brief exchange and is taken completely out of context. But that is part of the point. The positive and negative issues listed above are stated with respect to *local* patterns of argument. We identified each positive and negative point by looking at single sentences or, at most, single pairs of statements by Sud and Nord. But is this sufficient? Can we evaluate critical thinking performance by combing through a dialog (or an individual's thought process) statement by statement (or thought by thought) in search of virtues and vices?

There are reasons to think we cannot. First, we had a great deal of difficulty nailing down many of the positive and negative points listed above. In some cases, both positive and negative points seemed to apply to the *same aspects of the same statements*! For example:

- Are Sud and Nord to be *blamed* for discounting conflicting evidence offered by the other, or should they be *praised* for offering reasons in defense of their own positions and, in turn, challenging the other? Whether we blame Sud or Nord for presenting only evidence that confirms their own point of view may depend on whether they are really looking for the "truth," and are ultimately willing to change their minds if the other's argument is superior. Focusing on one side of a question may not be a fallacy in a dialogue context, where each participant has a different role to play – as long as it is clear that this is in fact the context. We need to know more about Sud and Nord and the implicit rules of the dialogue they have undertaken.
- Was Sud was too dogmatic in using the word "never"? The answer may depend on how Sud responds to further challenges.

- Were Sud and Nord obligated to make assumptions explicit? The answer may depend on how much knowledge of the domain each is entitled to assume the other has. It is probably impossible to enumerate all relevant assumptions, so it must be permissible to omit explicit statement of an assumption if the two participants share relevant background knowledge. On the other hand, implicit assumptions can lead to disastrous misunderstandings.

There is an even more important reason to question a local approach to evaluation. A whole new set of positive and negative issues comes to light when we look at the dialogue as a whole in its actual context. To do this, we must adopt an external perspective rather than an internal one. For example, from this broader point of view, the following issues emerge:

- Most importantly, what is the **real purpose** of answering the question (where will the enemy attack)? Is it **important to answer that question**? How will the answer support the mission? What is the overall planning context that makes this issue important? For example, if we are planning to attack the enemy first, it may not be all that important to determine the enemy's intent.
- Assuming this dialogue is worthwhile, it cannot go on forever. **When should the identification, challenge, and defense of assumptions stop?** How does that decision depend on the purposes of the dialogue?
- Is the issue of enemy artillery range **important enough** to warrant discussion? If it is important enough in this context, would it be important enough to discuss in all contexts? Have Sud and Nord gone to a **level of detail** on the single issue of artillery that may cause them to miss other important issues? Have they unduly sacrificed breadth for depth?
- What are the **common assumptions** that constrain the perspectives of both participants? For example, are they both assuming a particular enemy objective, and that the enemy will take the offensive to achieve it? Why has neither of them asked whether we can **influence** the enemy's action rather than simply trying to **predict** it?
- What **additional information** is available to shed light on the question, e.g., from the intelligence officer or from other units? Are Sud and Nord actively considering their information sources and options?

If questions like this are relevant to the evaluation of critical thinking, then it must be defined, understood, and trained as more than a correct *relationship among sentences or propositions*. In other words, it is more than logic, whether formal or informal. It must be viewed instead as a *process* carried out by real persons in a real context. The process has *internal* goals and constraints, one of which is to ask critical questions that bring to light relevant alternative possibilities. At the same time, the process has *external* objectives, beyond the process itself, which in this example concern the mission. External objectives determine the duration and depth of questioning, and the span of alternative possibilities to be considered. The participants should ask whether this dialogue is a reliable method for achieving those external objectives. Critical thinking is asking *questions* about alternative *possibilities* in order to achieve some *purpose*. Both its internal constraints and its real-world purpose must be included if critical thinking is to be properly evaluated and trained.

3. SCOPING THE TERRAIN

What Discipline Does Critical Thinking Belong to?

At the top level, we are interested in three basic questions (Figure 1): One is normative: *How should we think critically, i.e., what counts as good critical thinking?* The second is empirical (descriptive and/or explanatory): *How do people think critically, i.e., how is critical thinking actually accomplished?* And the third is applied: *How can critical thinking be improved?* Taken together, normative, empirical, and applied issues provide a set of systematically interconnected answers to the question, *What is critical thinking?* Each level must be considered in deciding what it is we should be trying to train, and how to train it.

Critical thinking has traditionally been approached from three distinguishable points of view, with different interests, assumptions, and methods of inquiry. Each of the three categories corresponds to a family of disciplines (subject, of course, to its own internal diversity of interests and methods):

1. The normative category includes areas of philosophy such as epistemology and formal logic, but also decision theory, informal logic, dialogue theory, communication studies, rhetoric, argumentation theory, artificial intelligence, forensics and debate, law, and critical studies. It also includes the implicit or explicit practices, canons, or standards that are applied in reasoning and argument within any specific discipline or science.
2. The descriptive/explanatory category draws on relevant work by cognitive psychologists, at either the process or mechanism level. Psychological research might be relevant because it directly addresses cognitive processes involved in problem solving, reasoning, decision making, creative thinking, and inference, or the processes of development and learning by means of which they are acquired. Psychological research may also be relevant because it addresses the underlying mechanisms by means of which those processes are executed, including perception, attention, long-term memory, working memory, affect, and knowledge representation. We refer to the empirical (descriptive/explanatory) category as cognitive. But in principle this category should be taken broadly, to include other descriptive and explanatory disciplines, such as social psychology, sociology, linguistics, discourse and conversation analysis, and speech and communication theory.
3. Finally, the applied category includes education, instructional theory, training, educational psychology, human factors engineering, expert systems, decision support and decision aiding, and the existing practices and knowledge of instructors and trainers.

These three areas of research can and have been pursued independently, but critical thinking brings them together. The links among them are a crucial part of any effort to unify the subject.

How Are the Top-Level Questions Connected?

Bell, Raiffa, and Tversky (1988) and others have distinguished between normative, descriptive, and prescriptive aspects of decision making. The normative aspect specifies how we *ought* to make choices. The descriptive aspect specifies how we *in fact* make choices. The prescriptive aspect is designed to bridge the gap: to assess the discrepancy between normative and descriptive and help real people make better choices. These categories closely parallel our

distinction among normative, cognitive, and applied issues. However, an important objective of the critical thinking framework is not simply to distinguish these three types of issues, but to articulate the connections among them. As shown in Figure 1, Bell, Raiffa, and Tversky (1988) emphasize the influence of both normative and descriptive considerations on prescriptive applications, but they see no interaction at all between normative and descriptive approaches. Moreover, each approach utilizes different criteria of validity:

> Descriptive models are evaluated by their empirical validity, that is, the extent to which they correspond to observed choices. Normative models are evaluated by their theoretical adequacy, that is, the degree to which they provide an acceptable idealized account of rational choices. Prescriptive [applied] models are evaluated by their pragmatic value, that is, their ability to help people make better decisions. (p. 8)

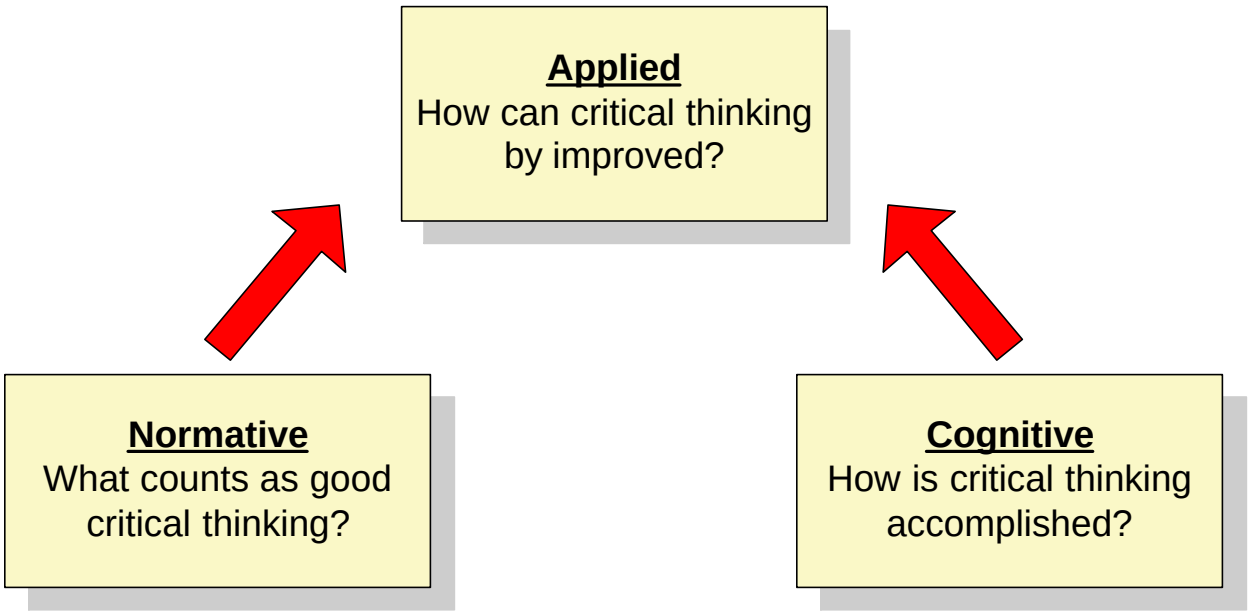


Figure 1. Three major categories of issues about the critical thinking. Arrows represent the standard view of how the three kinds of issues are related.

In contrast to Bell et al., our conception depicts a more tightly woven web of connections among the three levels (Figure 2). It embodies three explicit *hypotheses* about the connections among these three categories of issues:

1. Normative models have a significant influence on cognitive models.
2. Cognitive models have a significant influence on normative models.
3. Cognitive and normative considerations must be integrated in the development of applied tools for training and assessment.

The first hypothesis is based on David Marr's (1982) theory of levels of analysis of an information processing system. In this framework, a normative account specifies the objectives, constraints, and necessary functions of a process or faculty, while the cognitive theory describes (a) the mental representations and processes by means of which those functions are performed,

and (b) the cognitive mechanisms that underlie the representations and processes. The normative theory thus sets the target and boundary conditions and supplies high-level concepts and assumptions. It thereby clarifies the significance and improves the efficiency of the cognitive research.

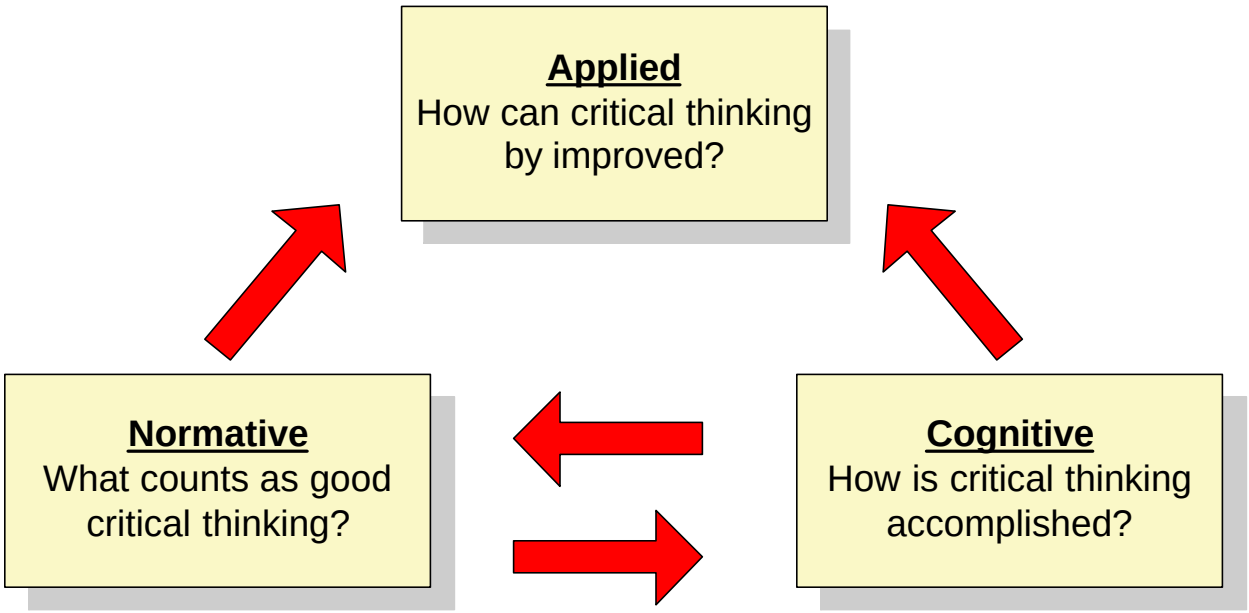


Figure 2. Arrows represent hypothesized interactions among the top-level issues regarding critical thinking.

Perhaps nothing divides researchers in both cognition and philosophy more than their disagreement on the second hypothesis, the influence (if any) of cognitive findings on normative theory. There are numerous gradations of views on this issue (Kim, 1994; Haack, 1993: pp. 118-138). At one extreme, is what we have called *analytically based* prescription (Lipshitz & Cohen, 2001), which aims at a ideal of rationality that is purportedly not influenced by how people actually think. This corresponds in philosophy to the position that normative principles are known *a priori*, i.e., independent of experience or empirical knowledge, including theory in cognitive science. The best examples, perhaps, are formal deductive logic and the probability calculus. At the other extreme, is the view that normative issues are dissolved or subsumed in empirical science; strictly normative concerns should be abandoned in favor of empirical investigation of how people actually think. This view overlooks the contribution of normative analysis of the purposes and functions of cognitive processes and faculties. It is like saying that mathematics is irrelevant to understanding a calculator, since we can explain what a calculator does by looking at its circuits. The most reasonable position in this debate, in our view, is what we call *empirically based prescription* (Lipshitz & Cohen, 2001), which allows and indeed *requires* that normative theory take actual thinking processes into account, while nonetheless recognizing the existence of a distinct set of normative concerns.[2] From this perspective, normative theory is not prior to or privileged over empirical science; normative and cognitive

[2] The view that there is a reciprocal influence between the cognitive on the normative is a moderate version of *naturalized epistemology* (Kornblith, 1994). This is distinct from, but consistent with, *Naturalistic Decision Making (NDM)* in psychology (Klein, 1993; Cohen, 1993a).

issues interpenetrate each other. There is really only *one* theory of critical thinking, with normative, cognitive, and applied poles.

For example, *argument* is a central concept in critical thinking, and the role that it has played illustrates the mutual interpenetration of normative and cognitive theory. Sound argument has traditionally been regarded (by both formal and informal logicians) as a necessary and sufficient condition for extending our knowledge by reasoning (as opposed to perception); i.e., sound argument permits us to accept new beliefs based on *inference* from beliefs that we already accept. Determining the soundness of an argument is supposed (at least by formal logicians) to be a matter of applying *context-independent* criteria to the sentences that make up the argument in question. These ideas have been directly imported from logic and philosophy into research by psychologists on reasoning and decision making, as well as into critical thinking instruction. The result is an inadequate normative paradigm and slower progress in both cognitive research and application. Neither context-independence nor the primacy of argument holds up against better understanding of real-world reasoning. A more sophisticated understanding suggests that beliefs are justified by their participation in a mutually supporting network of beliefs. The network can never be made completely explicit in terms of the premises and conclusions of an argument. Instead of describing self-contained valid inferences, argument is a tool for bringing a selected subset of those interconnections out into the open. Cognizers must determine how much of the underlying knowledge needs to be made explicit to fill gaps in information or clarify assumptions in a particular dialogue context. Argument needs to be put into a more appropriate perspective, as a situation-specific and resource-limited *strategy* for handling and sharing knowledge.

Few researchers disagree with the principle that both cognitive and normative issues are relevant to training critical thinking. But they do differ regarding the size of the gap between the way people ought to think and the way they actually think. The gap will be larger if normative and cognitive models are developed in relative independence of one another. For analytically based prescription, there is no positive role for the description of actual cognitive processes in normative modeling. Normative models provide a description of competence, while cognitive models deal with factors that degrade the translation of competence into performance. The role of cognitive modeling is therefore limited to negative findings, e.g., regarding biases in cognitive processes (Kahneman, Slovic, & Tversky, 1982) or the limited capacity of cognitive mechanisms (Payne, Bettman, & Johnson, 1993). At the applied level, training and decision aiding will try to fundamentally change the way people think, in ways that run against the grain of their actual propensities and abilities (Cohen, 1993b,c). For example, they may be asked to use explicit logical rules or to quantify their degrees of belief and preference. The *content* of training will be based on formal normative models, while only the *techniques* and *methods* used in training will be influenced by psychological findings regarding limits on performance. Examples of this approach to training thinking include Baron and Brown (1991) and Nisbett (1993).

On the other hand, the size of the gap between the normative and the cognitive will be smaller if the second and third hypotheses are accepted. From the empirically based perspective, cognitive modeling has a positive, direct influence on normative criteria. For example, psychologists who study expert performance (e.g., Ericsson & Smith, 1991) look for characteristics of reasoning that distinguish experts from novices and that are associated with a higher probability of success in real-world tasks. Dialogue theorists (e.g., Walton, 1998; van Eemeren, Grootendorst, Jackson, & Jacobs, 1993) observe actual argumentation as it is

conducted in different fields of everyday life, and attempt to characterize the purposes of the dialogues and of the norms imposed by participants on one another (Jackson, 1989; Johnson, 2000). From these observations, they derive idealized models of how such dialogues might best be conducted to achieve their purposes. From the empirically based point of view, training works with the cognitive processes that actually occur, and determines how to make them more effective. In this case, the *content* of the training, as well as techniques and methods, is influenced by psychological research. Examples of this approach include many chapters in Voss, Perkins, and Segal (1991) and Chipman, Segal, and Glaser (1985) as well as Cohen, Thompson, Adelman, Bresnick, Shastri, and Riedel, (2000b).

Paradoxically, a closer relationship between normative and empirical models makes the normative models more useful. Normative principles are useless unless they apply to *actual* cognitive processes and mechanisms and the real-world functions they serve. Normative and cognitive models will never be identical: Humans do make errors, sometimes quite systematic and serious ones. But such errors are best understood when normative and cognitive theories are interpreted as addressing the same purposes under similar constraints. Normative models and processes must be close enough to actual models and processes for the discrepancies to be of interest (Cohen, 1993a,b). Conversely, cognitive theories are useless without normative guidance. Just as the structure of the eye makes little sense unless its function in vision is understood, in an important sense cognitive theories cannot properly understand *what* they are studying without a grasp of normative purposes and constraints. Thus, a degree of convergence between normative and cognitive is not only necessary for effective training, it is also a source of validation for both of them. A goal of the framework presented in this report is to map out some directions in which such a mutual adjustment of normative and cognitive theories might proceed.

What Does It Mean to Define Critical Thinking?

There are many definitions of critical thinking, as we shall see. But very few authors prepare the ground by asking first what it *means* to define critical thinking.[3] Given the differences among normative, cognitive, and applied perspectives, the result is likely to be confusion and miscommunication. Before trying to define critical thinking, we need to ask (briefly) what approaches are available to us. What would it *mean* to know what critical thinking is? An adequate framework must accommodate complementary perspectives.

Consensus Based Definition

One approach is to "capture the central tendencies of our contemporary usage of the term" (Ennis, panel discussion, 20 March 2000), excluding the areas where there are major differences. Our review of definitions in the literature (see below) suggests that a common core does exist, even though it falls far short of exhausting everything that might be said. To anticipate our findings, we can safely regard critical thinking as, at least, *the deliberate evaluation of intellectual products in terms of an appropriate standard of adequacy*. We will also consider the way current definitions differ. The common core of usage leaves indeterminate what the products to be evaluated are (although they include *beliefs* at a minimum), by what standards they are evaluated (e.g., logic or something else), and what processes are used to arrive at a judgment (e.g., is it necessary to consider alternatives?). Ennis's own definition ("reasonable, reflective thinking that is focused on deciding what to believe or do") might do just

[3] Ennis is a major exception.

as well as a consensus definition. It differs from the starting point suggested above only in excluding other intellectual products, such as stories or art works, in addition to beliefs and actions.

Applied Definition

The consensus-based strategy is a useful start, but it cannot provide rigorous guidance for training. For example, Ennis (panel discussion, 20 March 2000l) distinguishes between his *concept* of critical thinking ("reasonable, reflective thinking that is focused on deciding what to believe or do") and his *conception*. While the concept is "loose and open-ended, leaving many things as yet not necessarily decided," the conception of critical thinking fills in these details with a list of component dispositions and abilities (Ennis, 1996). One version of Ennis' conception of critical thinking, i.e., his applied definition, is reprinted as Appendix A. Unfortunately, the connection between the concept of critical thinking and the more detailed conception is inevitably, and understandably, loose and informal. There is little systematic rationale in the concept itself for what to include and what not to include in the detailed conception, other than informed intuitions about desirable educational objectives.

Cognitive Mechanistic Definition

One response to this problem is resigned acceptance: At the start of an inquiry all that is needed is a loose, informal understanding of a phenomenon based on common usage, but a more definitive answer should come toward the end, rather than at the beginning, of scientific investigation. It will be the task of psychologists and other researchers to uncover the cognitive (and perhaps social) *mechanisms* underlying critical thinking, and knowledge of these mechanisms is necessary to tell us finally "what critical thinking is." Such a principled and theory-based approach will attempt to "carve nature at the joints" (Baron, panel discussion, 20 March 2000), rather than simply conform to current usage. It will provide a solid basis for rigorously specifying educational objectives.

The Need for a Better Approach

Unfortunately, as Ennis argues, a consensual definition model is not adequate for today's needs, because common usage is not specific enough to support the design of curricula. And educational applications cannot wait for the complete scientific understanding that is necessary for an eventual mechanistic definition. Perhaps more importantly, even if a detailed mechanistic account were available, it is not clear *how* it would help support practice in the absence of understanding at other levels. For applied purposes, detailed lists of skills and dispositions such as those provided by Ennis and others are more relevant. A variety of educational applications already exist based on such fleshed out conceptions.

In this section we have already alluded to three strategies for defining critical thinking:

- *Consensual*: a central tendency or consistent, common core of current usage; plus some understanding of the variations and inconsistencies in current views
- *Cognitive mechanistic*: the cognitive pulleys, levers, and gears underneath critical thinking, an account of which is the end product of empirical research and modeling
- *Applied*: interim but detailed guidance for education, training, and assessment

A natural question is: What can cognitive theory and research tell us *now* about critical thinking? How can we augment and refine the current concept to increase understanding, facilitate successful theory-building, and provide a more reliable basis for applied prescription?

Our framework identifies two additional strategies for clarifying the concept of critical thinking. Both of these strategies specify critical thinking in terms of general characteristics of a computational device, such as the human mind or brain. The first strategy is cognitive, but at a more abstract level than mechanism. It constrains critical thinking *processes*, in terms of the general class of transformations it performs on cognitively accessible internal representations. The second strategy is *normative*. It characterizes the transformations that critical thinking *should* perform on cognitively accessible internal representations in order to accomplish its goals. The normative strategy provides useful constraints on cognitive theory. Figure 3 summarizes all five strategies for defining critical thinking.[4]

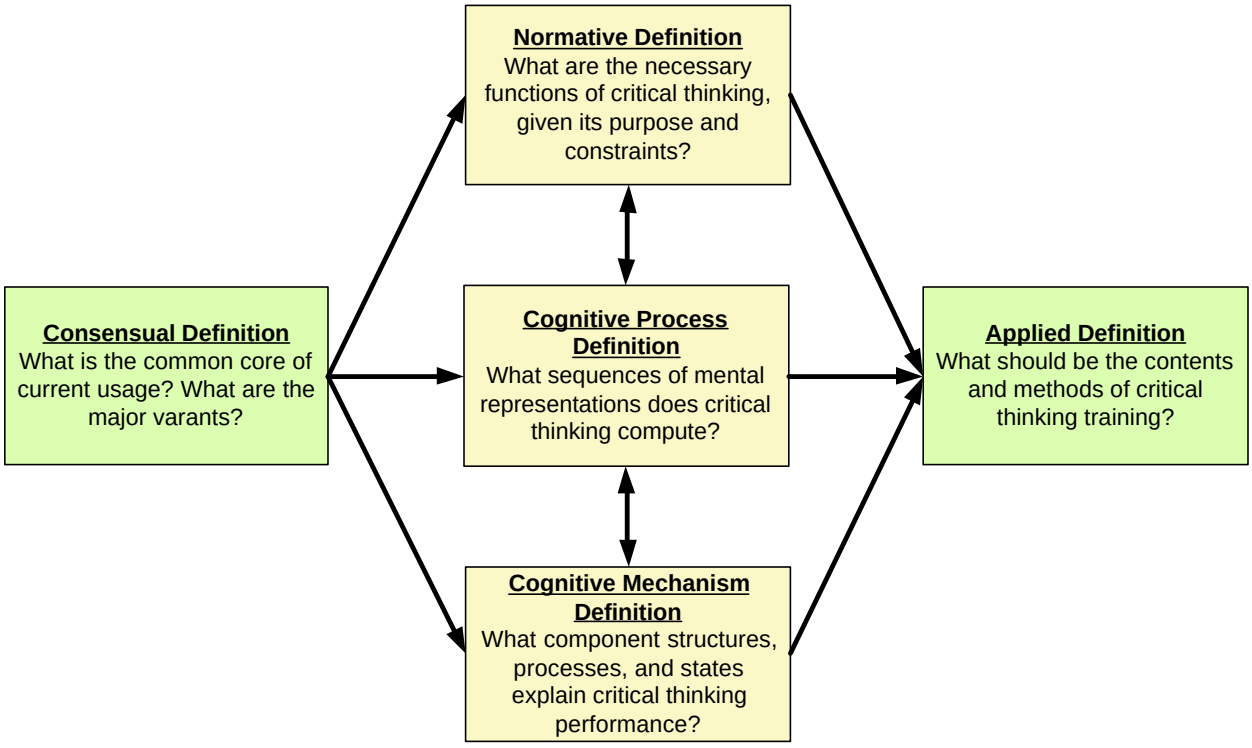


Figure 3. Five different answers to the question, *What is critical thinking?*

[4] We are referring to these as different types of *definition*, but there is nothing magical about that word. There is no hard-and-fast distinction within any particular theory between its *definition* of X (e.g., critical thinking) and other central statements about X that are asserted within the theory. All claims offered by theory must be taken together if the theory is to be understood, used, and evaluated. According to Quine, we don't need definitions to understand the meaning of a term (Quine, 1993; p. 198): "No definition has been given of 'electron' or 'neutrino'. Most theoretical terms in the sciences are introduced by description but not defined. The important thing about introducing the term is that it should help in systematizing and simplifying a theory whose test points lie in observation." A complete theory of critical thinking will include analyses, hence, "definitions," at several different levels.

How Are Normative and Cognitive Definitions Related?

Further insight comes by considering the classification of theory levels first proposed by Marr (1982), and alluded to briefly above. Marr studied vision, but his classification has more recently been applied to cognition in general, including causal inference and problem solving (Anderson, 1990) and reasoning (Johnson-Laird & Byrne, 1991; Oaksford & Chater, 1998). Marr distinguished three "levels at which an information processing device must be understood before one can be said to have understood it completely" (p. 24). The three levels are:

- *Computational level*: What is the goal of the computation, why is it appropriate, and what is the logic of the strategy by which it must be carried out? This level describes *what* must be computed, not the optional details of *how* it is computed.
- *Representation and algorithmic level*: How can this computational theory be implemented? In particular, what is the representation for the input and output, and what is the process for the transformation of inputs into outputs?
- *Hardware level*: How can the representation and process be realized physically, e.g., in the brain?

Two of these levels correspond directly to top-level parts of our critical thinking framework: The computational level comprises *normative* issues, and the representation and algorithm level comprises *cognitive* issues. It would be more accurate to refer to Marr's computational level as the *normative* level, since it characterizes the functions that a particular type of device (i.e., cognitive system) *must* perform in order to be adapted to its environment and achieve its goals.[5]

As Marr points out:

> …an algorithm [i.e., a process] is likely to be understood more readily by understanding the nature of the problem being solved than by examining the mechanism (and the hardware) in which it is embodied…. In a similar vein, trying to understand perception by studying only neurons is like trying to understand bird flight by studying only feathers: It just cannot be done. In order to understand bird flight, we have to understand aerodynamics; only then do the structure of feathers and the different shapes of birds' wings make sense… (pp. 27-28).

The same point holds for such higher-level cognitive processes as problem solving, reasoning, decision making, creative thinking, and critical thinking. By analyzing both the goals associated with these processes and the circumstances in which the goals must be achieved, we may be able

---

[5] A number of commentators agree in finding the term *computational* misleading. It is not actually the computations that are described at this level, but the *goals* of the computation and the constraints imposed by the *environment* on processes for achieving the goals. Marr's *computational* level of analysis corresponds closely to what Newell (1981) called the *knowledge* level, Anderson (1990) called the *rational* level, and Dennett (1978) referred to as the *intentional* level. All of these concepts appear to share a common normative component.

Several qualifications are relevant in understanding this notion: (i) Normative is meant in an instrumental not an ethical sense: i.e., what must be done to achieve a specified result. (ii) Anderson (1990, p. 29) emphasizes the centrality of the concept of *adaptation* at the rational level of analysis. But a successful adaptation is not necessarily *optimal* in any global sense, since it is constrained by the specific evolutionary or individual history of an organism, which makes certain traits and behaviors available as solutions and others not. (iii) Putnam (1994, chapter 22) demonstrates that we can never reflectively know a complete description of our normative competence, because of Gödel's incompleteness theorem . But Putnam does not show that an incomplete description would not be useful.

to draw conclusions about the functions that must be performed (to achieve the goals in those circumstances). In other words, the normative level characterizes the *problem* (goals and obstacles to achieving the goal) that a higher-level cognitive process must solve, and this characterization provides some general constraints on the solution. If successful, this in turn helps narrow the search for an adequate theory of cognitive processes and structures (i.e., Marr's representation and algorithmic level):

> If the mind is not a random set of mechanisms, but is structured to optimize its adaptation [within constraints], one can use the hypothesis of optimization to guide the search for a scientific theory. Otherwise, one has to rely on very weak methods to search a very large space of psychological hypotheses. (Anderson, 1990, p. 30)

There is reciprocity of influence between the normative and cognitive levels (Oaksford & Chater, 1998). Modeling between the two levels is highly iterative. The normative level spells out the goal of the computational device, along with environmental constraints on the achievement of the goal, and functions that are necessary to achieve the goal under those constraints (e.g., if it is a calculator, it should respond *4* to the query, *2+2=?*). However, an initial characterization at the cognitive level is necessary first, to identify goals the device has evolved or been designed to serve (e.g., is it a telephone or a calculator?). The cognitive level also suggests some of the processes and mechanisms that are in fact used to achieve the purpose and characteristics of the environments in which they operate. This information then informs and guides normative modeling (e.g., How do features of the environment constrain achievement of the goals? Which of the observed functions are necessary in the light of which constraints?). Additional research at the cognitive level is then guided by these results, as it fleshes out the information processes and structures that are in fact used to compute the normative solution to the problem.

The representation-and-algorithmic level of Marr's classification correspond to the cognitive part of our framework. Anderson (1990) further subdivided it into two parts: the cognitive *process* level and the cognitive *mechanism* level, respectively. A cognitive process is analyzed in terms of *functions* that compute *sequences of mental representations* which correspond potentially to behavior (Anderson, p. 20). In other words, it is a theory of the input-output processes that generate *cognitively accessible contents*.[6] A theory at the process level can take the form of a program that predicts the general features and/or details of think-aloud protocols, in which individuals describe what they are attending to as they perform a cognitive task. Typically, theories of problem solving, decision making, and reasoning are at the cognitive process level. For example, a description of the problem space together with a flow chart for exploring it would serve as the process model of a problem-solving strategy (Newell, 1990). The cognitive mechanism level, on the other hand, explains how the computations specified at the process level are performed below the level of potential overt behavior or verbalization. It analyzes a cognitive process in terms of lower-level cognitive *structures* (such as working memory and long-term memory), lower-level *processes* (such as spreading activation, attentional filtering, and symbol writing), and hypothesized *states* or *traits* (such as learned rules or patterns,

---

[6] Anderson actually refers to this as the *algorithmic* level. We refer to it as the process level in order to avoid connotations associated with the algorithm / heuristic distinction, and also to permit less formal, non-algorithmic theories at this level.

schemata, I.Q., introversion-extroversion, and open-mindedness). Eventually, cognitive mechanisms may be specified in enough detail to make contact with the *hardware* level, where critical thinking might eventually be understood in terms of actual neural structures and processes in the brain.

With these distinctions in hand, we can identify two promising strategies for going beyond a consensual definition of critical thinking without requiring a complete cognitive mechanistic model. One strategy is normative, the other is based on cognitive process.

Normative Definition

To understand cognitive processes and mechanisms, we must know *what* they are designed (or have evolved) to compute: i.e., what the purpose is, and *how* it must be accomplished given various *facts* about the environment and the device itself. Here is a template for such a normative definition:

Template for a normative definition of a cognitive process and/or cognitive mechanism

Purpose	*What the process or mechanism is supposed to compute.*
Constraints	*Facts that influence the way the purpose can or cannot be achieved within the environments in which the process or mechanism will operate.*
Function	*Functions the process or mechanism must perform in order to achieve the purpose within the constraints.*

Normative theory provides an abstract characterization of the *adaptive functions* which critical thinking must perform to be successful, and thus provides essential constraints on a cognitive model of critical thinking. Normative theory may help also us *define* critical thinking at the cognitive level by showing how its adaptive functions differ from the adaptive functions associated with *other* high-level cognitive processes such as problem solving and reasoning. We will consider strengths and weaknesses of different normative definitions in more detail later.

Cognitive Definitions

Figure 4 outlines a cognitive framework that will help lay the groundwork for a cognitive definition and theory of critical thinking. It has two dimensions: *internal / external* from top to bottom, and *time span* from left to right.

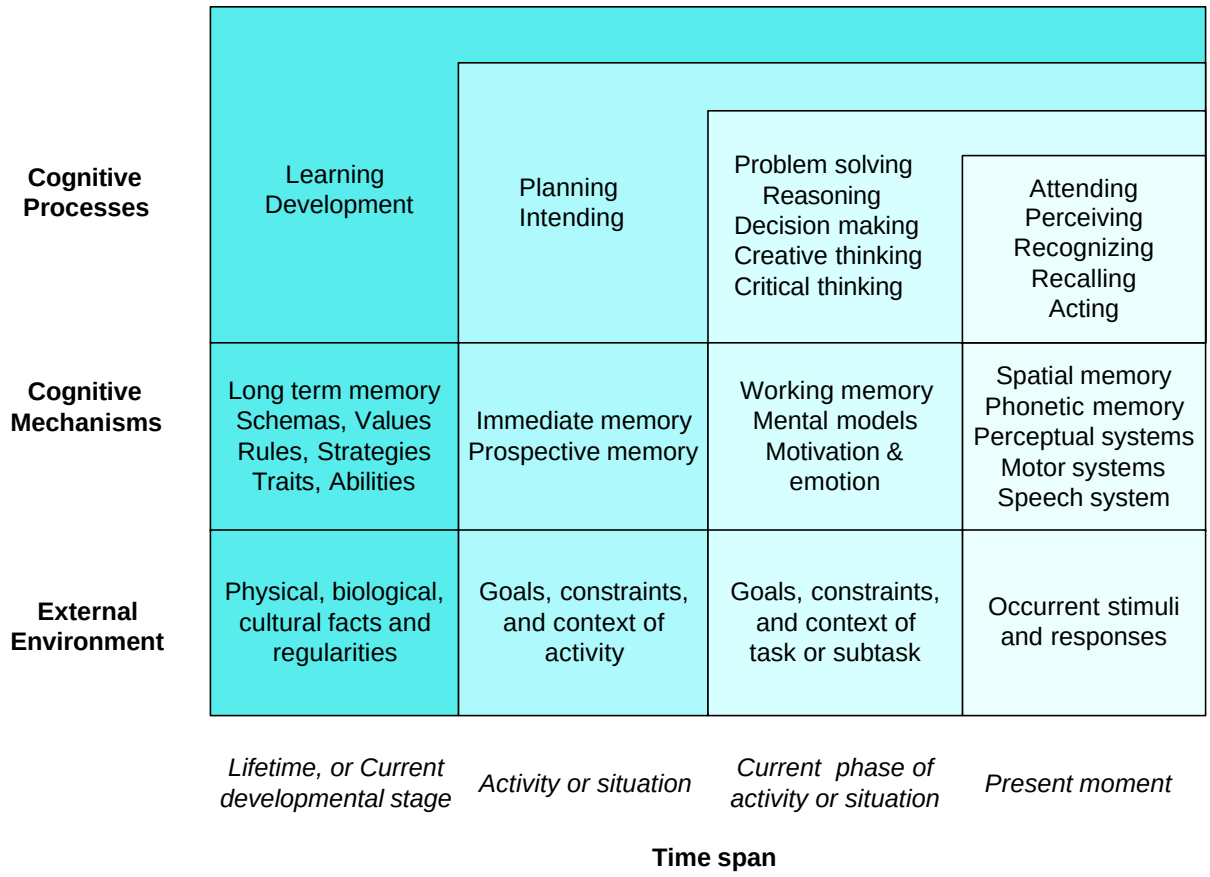


Figure 4. A schematic cognitive framework showing relationships among cognitive processes and mechanisms. Time span is represented from left to right, and internality/externality from top to bottom.

Internal / external. From top to bottom, Figure 4 distinguishes three broad categories of phenomena: cognitive processes, cognitive mechanisms, and the environment. A cognitive process is an input-output function that generates mental representations that actively or potentially influence behavior (Anderson, 1990; pp. 20-21). The cognitive events thus generated (perceptions, thoughts, emotions, and sensations) are *internal* in the strongest sense, i.e., consciously accessible to the cognizer. "Cognitive process" may refer either to a specific actual occurrence of such events or to an abstract input-output function that predicts them. Strategies for learning, problem solving, decision making, and critical thinking are processes in the latter sense. They are accessible to the cognizer via the contents they produce on specific occasions.[7]

Cognitive mechanisms are devices that implement the relevant input-output functions or programs. Mechanisms like working memory, attention, schemata, skills, personality traits, and motor systems are not known by introspection, but are postulated by psychologists to explain observed behavior, including reportable conscious processes. A particular process may require a combination of mechanisms, such as attention, working memory, and long-term memory, in

[7] See Astington, Harris, and Olson (1988) for studies of how people develop theories of their own and others' minds.

order to "run." Performance capabilities as well as limitations in executing the process are accounted for by the combination of theoretical parameters and resource constraints associated with the relevant mechanisms.[8]

Time span. From left to right, Figure 4 classifies cognitive processes and mechanisms according to the time span over which they tend to operate or, equivalently, the rate at which their contents or features change. Combinations of cognitive mechanisms and processes operate over different time spans. Each vertical slice represents a set of loops that are embedded within the loops to its left. For example, learning processes, long term memory contents, values, individual abilities, and traits change slowly and are sustained over relatively long time scales, from events in the distant past to the present. Immediate memory for recent events and prospective memory for intended actions operate over shorter time scales, such as the span of a coherent activity or project. Decision making, problem solving, reasoning, critical thinking, and creative thinking involve the construction and transformation of mental models in working memory. These processes and contents persist for short periods at appropriate phases within an activity or task. Input processes (e.g., attending and perceiving the contents of sensory memories) change rapidly and persist over briefer time spans. Actions executed by means of motor systems also occur in real time, in parallel with input processes.

Figure 4 does not show the flow of information and control by the ordinary conventions, e.g., by arrows directly linking long-term memory and perceptual inputs, respectively, to working memory, and working memory to action. Nor does it show feedback loops, e.g., from the outcome of action back to long-term memory. This information is represented by a stipulation: Information stored in *any* cognitive mechanism is available to *all* processes and mechanisms embedded within the span of operation of that mechanism (i.e., anywhere at the same time slice or to its right in the diagram). A longer time-scale mechanism thus provides the context for each shorter-time scale process. For example, personality traits as well as the contents of long-term memory may influence processes of planning, problem solving, and attending; and the contents of immediate memory and prospective memory are available throughout cycles of perceiving, decision making, and acting. Shorter time scale processes may iterate as often as necessary within the span of longer time scale processes. Conversely, the shorter time-span processes *may* return new information to any of the processes in which they are embedded or influence its future operations in other ways. For example, processes of learning and planning may utilize information computed by sub-processes of problem solving, decision making, and critical and creative thinking, while the latter may utilize information computed by sub-processes of attending, perceiving, and acting. In this way, relatively fleeting information helps modify more enduring structures and processes.

---

[8] Cognitive mechanisms are more "external" than cognitive processes because they are less directly associated with sequences of consciously accessible contents. Nevertheless, accessibility is a matter of degree, since mechanisms help explain, and thus are *inferred* from, conscious contents and behavior. For example, a cognizer may infer her beliefs and values from occurrent feelings and thoughts. Since schemas, values, or beliefs influence the contents of cognitive processes, long-term memory can be thought of as (very loosely speaking) having cognitively accessible contents. Schemas, values, and beliefs are, nevertheless, part of the long-term memory *mechanism*, because for example they incorporate theoretical assumptions about the way information is organized for storage and retrieval under resource constraints.

Where Does Critical Thinking Fit in?

We have located critical thinking on each of the two dimensions of Figure 4. Critical thinking is a *process*, in the sense of an input-output function or program, with a time span corresponding to a *particular phase of an activity or situation.* This is not a complete definition because in both respects it resembles problem solving, reasoning, decision making, and creative thinking. Each of these processes performs transformations on mental representations in working memory. They are all supported or influenced by mechanisms of working memory, mental models of the situation, emotions, and motivation. They all occur within the context of knowledge stored in long-term memory and plans stored in prospective memory. They typically span a particular phase of an activity or situation, and may be enlisted as sub-processes for longer-range learning or planning. External input or output are not a necessary part of any of them, although attending, perceiving, and acting may be enlisted as sub-processes when needed.

A more complete cognitive model of critical thinking must *differentiate* it in a systematic way from the other cognitive processes that operate over the same time span and involve transformation processes supported by the same cognitive mechanisms (e.g., problem solving, reasoning, decision making, and creative thinking). To be systematic, a process model must characterize critical thinking within the framework of a general computational theory, using the basic terms of that theory (Johnson-Laird, 1988, 1993). (Here we are using *computational* not in Marr's sense (to mean normative), but in the sense of a general theory of information-processing devices at the cognitive process level.) The theory will not only characterize the higher-level sequences of mental representations underlying critical thinking, but also show how they are similar to and/or different from related higher-level sequences of representations involved in problem solving, reasoning, decision making, creative thinking, and so on.[9] Thus, a promising initial approach to improving the definition of critical thinking is to see how it can be distinguished from other cognitive processes in general computational or information-processing terms.

Can critical thinking also be defined in terms of cognitive mechanisms? To get a better understanding of cognitive mechanistic models, it will be useful to introduce an additional concept that links processes and mechanisms. Although a single process may be the most salient event in accomplishing a particular normative purpose, typically such a process is the tip of an iceberg represented by a cognitive faculty. A *cognitive faculty* combines multiple mechanisms and processes operating over different time spans to support a normative purpose. Thus, while critical thinking can be identified for many purposes with a single type of process, if we pull back a bit, we see a bigger picture. From this broader perspective, critical thinking is a faculty that involves the coordination in a particular way of individual traits, learning, attention, working memory resources, stored knowledge, strategies for long-term memory recall, perception, and action.

A mechanistic model must do more than list the cognitive mechanisms that critical thinking draws on. It must distinguish critical thinking from problem solving, reasoning, decision making, and creative thinking in terms of how they use a highly overlapping set of mechanisms. Such distinctions can be made at the mechanism level only by reference to the cognitive processes that orchestrate the action of the cognitive mechanisms across different time spans.

[9] It might also turn out that the traditional paradigms of problem-solving, decision making, creative reasoning, and so on do not represent a useful taxonomy.

Thus, a mechanistic definition of critical thinking must be a component in a more comprehensive theory of critical thinking as a cognitive faculty.

In sum, we have identified three perspectives on critical thinking:

1. A *normative theory of critical thinking* will try to say what critical thinking is by specifying purposes, constraints, and adaptive functions that apply when people are thinking critically and not otherwise.
2. A *cognitive process theory* will try to say what critical thinking is by identifying sequences of internal representations and transformations that distinguish critical thinking from other cognitive processes such as problem solving and decision making, and which tend to achieve the normative goals of critical thinking within the normative constraints.
3. A *cognitive faculty theory* will try to say what critical thinking is by specifying how processes and mechanisms combine in critical thinking, i.e., how individual traits, learning, attention, working memory, long-term memory, consciousness, and motivation combine to produce processes that we recognize as examples of critical thinking. A *mechanistic* theory of critical thinking is the component of a cognitive faculty theory that explains how mechanisms are used to support relevant processes.

4. POSITIONS OF THE OPPOSING FORCES

How Is Critical Thinking Defined in the Literature?

To get started with either a cognitive or normative definition, we need a preliminary fix on what does and does not count as critical thinking. Current usage (i.e., a consensual definition) is not the only available guide (there is also historical usage, as well as mere stipulation), but it is a useful anchor, which ensures relevance to the on-going debate. But there is an immediate problem. The critical thinking movement is a pragmatic answer to a perceived gap in current education. It aims to foster skills of independent thinking in a variety of different contexts, including school (elementary, secondary, college, and beyond; e.g., Wooditsch, 1991), work (Mitroff, 1998), and the general populace (e.g., Brookfield, 1987). In this pragmatic spirit, critical thinking content tends to be a hybrid, with no single consistent theoretical framework shared among various writers and researchers or indeed among the methods recommended by a single writer. Instead, researchers tend to draw in different ways and to varying degrees from disciplines such as logic, decision theory, rhetoric, philosophy, linguistics, psychology, and common sense. Are there, beneath the surface chaos of the definitions in the literature, important consistencies in the underlying concepts? Are there interesting and important reasons for the variations among definitions that are found?

To some degree at least, the answer is *yes*. One of the reasons that definitions diverge is that authors choose to address concerns at different levels, e.g., normative, cognitive process, or cognitive faculty / mechanism. Other divergences occur because authors address the same level at different scales of granularity or resolution. More fundamental differences arise when authors have adopted, sometimes implicitly, different views of the purposes, constraints, or necessary functions with which critical thinking must operate (We will address some of these more fundamental differences in the next chapter). Despite the variety of disciplines involved, our aim is to identify crucial themes as well as significant differences among definitions in the literature. To this end, the discussion is organized by categories – normative purpose, normative necessary functions, cognitive processes, and cognitive mechanisms – and by themes within those categories.

Normative Definitions

Purpose

Probably the earliest modern definition of critical thinking was proposed by Dewey (1910/1991). Although he referred to *reflective* rather than critical thinking, he explicitly contrasted the notion with *uncritical thought*. Dewey defined reflective thinking as:

> 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.

Many subsequent definitions echo Dewey's concept in part or whole. From a normative point of view, Dewey is identifying what we would call a *necessary function* of critical thinking – to think about the grounds and further implications of a belief. But Dewey is, at least implicitly, referring to a *purpose* of critical thinking as well, in the phrase "careful consideration of any belief." The first group of authors to be considered omits mention of the function of critical

thinking (looking at the grounds of a belief and further conclusions), but they refer more explicitly to its purpose: *assessment* of a *belief* or claim, in order to *decide whether to accept or reject it*: [10]

Author	**Purpose of CT is accepting or rejecting a belief**
Dewey (1910/1991)	Active, persistent, and careful consideration of any belief or supposed form of knowledge …*
Ennis (1962; quoted in Walters, 1994)	The correct assessment of statements.
Ennis (1987)	Reasonable, reflective thinking that is focused on deciding what to believe or do.
Moore & Parker (1998)	The careful, deliberate determination of whether we should accept, reject, or suspend judgment about a claim – and of the degree of confidence with which we accept or reject it.
Epstein (1999)	Evaluating whether we should be convinced that some claim is true … (p. 5)

Not just any acceptance or rejection will do, however. Many definitions make it more explicit that the purpose of critical thinking is to make *rationally acceptable decisions* about whether to accept or reject a belief or claim:

Author	**Purpose of CT is normatively correct acceptance of beliefs**
Ennis (1962; quoted in Walters, 1994)	The correct assessment of statements.
Ennis (1987)	Reasonable, reflective thinking that is focused on deciding what to believe or do.
Facione/American Philosophical Association (1990)	* While not synonymous with good thinking, critical thinking is a pervasive and self-rectifying human phenomenon.*
Wade & Tavris (1993)	The ability and willingness to assess claims and make objective judgments on the basis of well-supported reasons.
Siegel (1997)	Being a critical thinker requires basing one's beliefs and actions on reasons; it involves committing oneself to the dictates of rationality... (pp. 13-14) *

[10] To make the presentation simpler, in some cases we have split a single definition into parts which are displayed separately under different categories. An asterisk (*) signals that we have separated the parts of a definition, and that other components of that author's views appear under another category.

Epstein (1999)	Evaluating whether we should be convinced that some claim is true or some argument is good, as well as formulating good arguments. (p. 5)

The focus on beliefs as the targets of evaluation is inherited from the intellectualist tradition. For example, *logical* criteria can be applied only to beliefs, not to actions or other intellectual products such as designs, plans, stories, or paintings. Some authors expand the focus to include not only beliefs (considered as abstract propositions) but the actual *processes* that produce those beliefs as products. Evaluation of the thinking process may be undertaken to improve long-term thinking skill (i.e., a form of self-teaching), or it may be a means to a more rational decision about a particular claim. Self-reflection, like the focus on beliefs, was a component of the intellectualist tradition.

Author	**Purpose of CT includes assessment of thinking processes**
Oscanyan (1984)	… critical thought consists of the evaluation of mental acts…
Paul (1993)	…Thinking about your thinking while you're thinking in order to make your thinking better (p. 91)
Halpern (1996)	…Critical thinking also involves evaluating the thinking process – the reasoning that went into the conclusion…*
Levy (1997)	Think about thinking…the way we think…how well we think…why we think (p. vii)

Some definitions extend the *purpose* of critical thinking to consideration of *intellectual products other than beliefs*. The most limited extension of scope is to include actions as well as beliefs (e.g., Ennis and Halpern). This removes the traditional philosophical barrier between intellectual, cognitive products (beliefs) and emotive, motivational products (actions), i.e., mind and passion. Modern theory, of course, regards both beliefs and decisions as products of cognitive processes and states – although much of the critical thinking literature has not caught up with this. An even more dramatic extension of scope is to all products of the mind, or to any process of thinking regardless of what its product might be.

Author	**Purpose of CT includes consideration of targets other than beliefs**
Dewey (1910/1991)	Active, persistent, and careful consideration of any belief or supposed form of knowledge in the light of the grounds that support it…*
Glaser (1941; quoted in Walters, 1994)	…an attitude of being disposed to consider in a thoughtful way the problems and subjects that come within the range of one's experience…*
Reeder (cited in Govier, 1987, pp. 238-9)	As a critical thinker, one does not just let situations and claims slip by. Rather, one focuses upon and assesses beliefs, claims, events, discoveries, etc.*
Ennis (1987)	Reasonable, reflective thinking that is focused on deciding what to believe or do

Govier (1987)	Thinking about another product of thought (an argument, claim, theory, definition, hypothesis, question, or problem) in a special skeptically deliberative, evaluative way (p. 238)…. The product of critical thinking may be a well-formulated question, an improved definition, a second version of a poem, a new fashion design, or a better disposable diaper. It need not be the analysis of an argument (p. 240).
Johnson (1992/1996)	The focus of the critical thinker's scrutiny is *thought*… in its widest sense of being an intellectual/rational product of some sort, including such various items as beliefs, theories, hypotheses, new stories, and arguments, whether they are someone else's or one's own…(p. 225)
McPeck (1994)	Both the disposition (or propensity) and the relevant knowledge and skills to engage in an activity with reflective skepticism (p. 103)
Fisher & Scriven (1997)	Skilled and active interpretation and evaluation of observations and communications, information and argumentation (p. 21)

Necessary Functions

A large group of authors follows Dewey in elaborating on the functions that critical thinking must perform in order to achieve its purpose. Most of these authors confine critical thinking to the evaluation of beliefs or, perhaps, beliefs and actions – perhaps because it is easier to specify necessary functions of critical thinking if its purpose is defined narrowly, rather than more generally as the evaluation of any product of mind. These authors require that assessment include an explicit examination of the *reasons*, *grounds*, or *arguments* for a belief (or action).

Author	**Necessary functions of CT include assessing reasons for a belief**
Dewey (1910/1991)	Active, persistent, and careful consideration of any belief or supposed form of knowledge in the light of the grounds that support it…*
Facione/American Philosophical Association (1990)	We understand critical thinking to be purposeful self-regulatory judgment…*, as well as explanation of the evidential…* considerations upon which that judgment is based. *
Wade & Tavris (1993)	The ability and willingness to assess claims and make objective judgments on the basis of well-supported reasons.
Halpern (1996)	…Critical thinking also involves evaluating the thinking process – the reasoning that went into the conclusion we've arrived at or the kinds of factors considered in making a decision.
Siegel (1997)	Being a critical thinker requires basing one's beliefs and actions on reasons; it involves committing oneself to the dictates of rationality... To be a critical thinker one must be able, at least, to evaluate the evidential or probative force of reasons… (pp. 13-14)
Epstein (1999)	Evaluating whether we should be convinced that some claim is true or some argument is good, as well as formulating good arguments. (p. 5)

Some authors go into more detail, specifying that the evaluation of beliefs or actions must be in terms of *standards*, *criteria*, *general principles*, or an appropriate *basis*, but without explicitly requiring that the standard be logical:

Author	**Necessary functions of CT include use of criteria, not necessarily logical**
Oscanyan (1984)	On those occasions when a mental act is called into question, there is typically a need for some basis beyond the act itself for judging it. Here is where critical thinking comes to life: critical thought consists of the evaluation of mental acts, and concern about critical thinking involves identifying proper basis for evaluation and means for doing so.
Watanabe Dauer (1989)	The art of assessing truth claims according to certain general principles or canons.
Facione/American Philosophical Association (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. *
Lipman (1991)	…thinking that (1) facilitates judgment because it (2) relies upon criteria….* (p. 116)
Johnson (1992/1996)	The articulated judgment of an intellectual product … in terms of appropriate standards or criteria. (p. 226)
Paul (1993)	A unique kind of purposeful thinking in which the thinker systematically and habitually imposes criteria and intellectual standards upon the thinking, taking charge of the construction of thinking, guiding the construction of the thinking according to the standards, and assessing the effectiveness of the thinking according to the purpose, the criteria, and standards. (p.21)
Siegel (1997: p. 20)	Critical thinking demands a rejection of relativism. If we think there is some point to helping students become critical thinkers, we must think there are criteria, binding upon all reasoners, in accordance with which the strengths of reasons and arguments are appropriately determined, and we must think it is a good thing for students to master and utilize those criteria.
Diestler (1998)	A critical thinker is someone who uses specific criteria to evaluate reasoning and make decisions.

Some definitions are even more specific, asserting or suggesting that assessment must conform to an explicitly *logical* standard (in some sense of "logic"). Since formal logic applies only to *beliefs*, this necessary function narrows the purpose of critical thinking to the evaluation of beliefs, rather than other products:

Author	Necessary functions of CT include use of explicitly logical criteria
Glaser (1941; quoted in Walters, 1994)	*…knowledge of the methods of logical inquiry and reasoning, and some skill in applying those methods."
Flew (1998)	…thinking about thinking is concerned, at least in the first instance, with the validity or invalidity of arguments…arguments are concerned with the logical relations between propositions.
Freeley & Steinberg (2000)	The ability to analyze, criticize, and advocate ideas; to reason inductively and deductively; and to reach factual or judgmental conclusions based on sound inferences drawn from unambiguous statements of knowledge or belief. (p. 2)

A necessary function singled out by some authors involves *adjusting the standards* applied in critical thinking to *different contexts or tasks*. These authors are not all relativists, who deny the existence of general normative standards. Relativity to context might simply mean that application of a *general* standard (e.g., logic) requires careful interpretation of the situation, e.g., to determine what logical structure is intended, what premises are implicit, and what the words are intended to mean. Another mild form of context-dependence might occur because the general standard itself refers to contextual variables (e.g., criteria are adjusted to reflect the cost of errors or the time available for decision making). The latter would rule out logical criteria that refer only to the propositions being evaluated and omit facts about the context. It is often unclear which type of relativity is intended.

Author	Necessary functions of CT include adjusting to context
Paul (1993)	…disciplined, self-directed thinking which exemplifies the perfections of thinking appropriate to a particular mode or domain of thinking (p. 136)… coming up to these standards is relative and often has to be adjusted to a particular domain of thought…
Facione/American Philosophical Association (1990)	… explanation of the … contextual considerations upon which that judgment is based. *
Lipman (1991)	thinking that (1) facilitates judgment because it…(4) is sensitive to context (p. 116),
Halpern (1996)	…the thinker is using skills that are thoughtful and effective for the particular context and type of thinking task…. (p. 5)

Some authors make even stronger claims about the relativity of normative standards. They argue that there are no general standards of correct thinking. The standards themselves are specific to different domains:

Author	Necessary functions of CT include use of distinct criteria across domains
McPeck (1994)	… the various 'forms of thought'… have a logic, texture, and relevant background knowledge that are peculiar to themselves (p. 103)…Not only are canons of validity different, but what might be fallacious reasoning in one context or domain, might be perfectly correct in another (p. 109)

Cognitive Definitions

Cognitive Process Requirements

A cognitive process model should specify how the necessary functions of critical thinking are implemented. It describes a cognitive process as a sequences of cognitively accessible mental events. Definitions at the cognitive process level may be offered at different scales of granularity. They may specify processes in detail, e.g., in the form of computer programs. On the other hand, they may simply characterize the kinds of processes that are involved in a general way. At the most general level, cognitive processes blur into necessary functions.[11] Cognitive processes are cognitive *strategies* when they involve purposeful adaptation by an individual to a specific type of task environment. Strategies for evaluating the reasons for a claim differ in the factors they adapt to.

Some cognitive definitions describe processes that focus on one's own reasons for accepting a claim. These processes include identification of the *implications* of the belief, explicit seeking out of *reasons against*, identification and challenging *assumptions* underlying acceptance of the belief, *self-correction* of the reasoning process, challenging the *interests* that might have motivated the belief, and asking oneself critical *questions*. They may also include exposure of one's views to rebuttal by *articulating* them to others

Author	Cognitive processes include reflection on one's own position
Dewey (1910/1991)	Active, persistent, and careful consideration of any belief or supposed form of knowledge in the light of the grounds that support it, and the conclusions to which it leads. …certain subprocesses which are involved in every reflective operation. These are (a) a state of perplexity, hesitation, doubt; and (b) an act of search or investigation directed toward bringing to light further facts which serve to corroborate or to nullify the suggested belief. (p. 9)
Brookfield (1987)	Identifying and challenging assumptions…challenging the importance of context…* (pp.7-9)
Paul (1993)	… It comes in two forms. If the thinking is disciplined to serve the interests of a particular individual or group, to the exclusion of other relevant persons

[11] In any case, the divide between necessary functions and cognitive processes that implement the functions is not very sharp. The same sequence of events might be both essential for achieving the purpose of critical thinking (hence, a necessary function) and a consciously accessible strategy described at a coarse level (hence, a cognitive process). The distinction is clearer if normative models ignore the way reasoning is actually conducted, e.g., we limit necessary functions to the application of logical criteria.

	and groups, I call it *sophistic* or *weak sense* critical thinking….* (pp. 137-138)
Facione/American Philosophical Association (1990)	* …CT is a pervasive and self-rectifying human phenomenon… The ideal critical thinker is… 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.*
Lipman (1991)	Skillful, responsible thinking that facilitates good judgment because it… (2) is self-correcting…
Johnson (1992/1996)	The articulated judgment of an intellectual product arrived at on the basis of plus-minus considerations of the product…. (p. 226)
Browne & Keeley (1998)	1. Awareness of a set of interrelated critical questions. 2. ability to ask and answer critical questions at appropriate times…* (p. 2)

Another group of cognitive definitions adopts a broader perspective. It does not focus simply on finding and correcting mistakes in one's own beliefs, but on the active consideration of *alternative hypotheses and points of view*. It describes processes in which the focus of attention shifts to other peoples' reasons for accepting a contrary claim. These strategies bring other parties into the dialogue, at least hypothetically, as active proponents rather than simply as critics of one's own position:

Author	**Cognitive processes include reflection on alternative positions**
Brookfield (1987)	*… try to imagine and explore alternatives… reflective skepticism. (pp.7-9)
Paul (1993)	… * If the thinking is disciplined to take into account the interests of diverse persons or groups, I call it *fairminded* or *strong sense* critical thinking. (pp. 137-138)
Facione/American Philosophical Association (1990)	* The ideal critical thinker is habitually inquisitive, well-informed, trustful of reason, open-minded, flexible, fair-minded in evaluation… *
Missimer (1994)	A reasoned judgment, which must take account of other reasoned judgment(s) on an issue (p. 119)
Walters (1994)	Exercises reflective autonomy in her responses to competing ideas from both the intellectual and political marketplaces… ability to weigh particular claims against the background of broader concerns and alternative perspectives. (p. 18)

Cognitive Mechanism Requirements

Few definitions of critical thinking spell out in any detail the contributions of specific cognitive mechanisms. Nevertheless, roles of specific mechanisms are often implicit in the way

cognitive processes, or strategies, are described. For example, if a cognitive process is characterized as *thoughtful* or *reflective*, focal attention and consciousness are implied. If the process is described as *deliberate,* conscious control must also be involved.

Author	**Critical thinking must be under conscious control**
Dewey (1910/1991)	Active, persistent, and careful consideration of any belief or supposed form of knowledge …*
Glaser (1941; quoted in Walters, 1994)	…an attitude of being disposed to consider in a thoughtful way the problems and subjects that come within the range of one's experience…
Reeder (cited in Govier, 1987, pp. 238-9)	Critical thinking involves a reflective attitude. As a critical thinker, one does not just let situations and claims slip by. Rather, one focuses upon and assesses beliefs, claims, events, discoveries, etc. This focusing is not adventitious, but results from a conscious decision to think about or think through the things one encounters, and to develop habits which promote the implementation of such a decision.
Ennis (1987)	Reasonable, reflective thinking that is focused on deciding what to believe or do
Facione/American Philosophical Association (1990)	We understand critical thinking to be purposeful, self-regulatory judgment… *
Lipman (1991)	Skillful, responsible thinking that facilitates good judgment because it… (2) is self-correcting…
McPeck (1994)	Both the disposition (or propensity) and the relevant knowledge and skills to engage in an activity with reflective skepticism
Moore & Parker (1998)	The careful, deliberate determination of whether we should accept, reject, or suspend judgment about a claim….
Browne & Keeley (1998)	1. Awareness of a set of interrelated critical questions. 2. ability to ask and answer critical questions…* (p. 2)

Some authors require not only that the process of critical thinking be self-aware, but that it be guided by a *self-concept* of the thinker as the active shaper of his or her own thoughts:

Author	**Critical thinking requires a self-concept as active shaper of thought**
Paul (1993)	disciplined, self-directed thinking …
Paul (1993)	… taking charge of the construction of thinking…(p.21)

Some authors assert that critical thinking is *effortful*, i.e., that it draws heavily on cognitive capacity, because it involves overcoming strong pre-existing tendencies. As a result, affective or emotive mechanisms may also be involved.

Author	**Critical thinking is effortful or unpleasant**
Dewey (1910/1991)	* Reflective thinking is always more or less troublesome because it involves overcoming the inertia that inclines one to accept suggestions at their face value… (p. 13)
S. Fisher & Spiker (2000)	…we consider CT to be truly time-limited, wherein an individual may execute the necessary skilled processes for only a few minutes (or less) before he must "come out"… we assume that engaging in CT processes has true state-like consequences in which the individuals experience emotions, motivations, and other phenomenological experiences that are reportable…the experienced consequences of being in the state are generally unpleasant. (p. vi)

Many definitions not only refer to critical thinking as an activity or set of skills, but also require a persisting state, such as a *critical attitude* or an *actively open-minded disposition* to perform the activity. Some definitions suggest a rather weak interpretation of this state. The required attitudes and dispositions are nothing more than tendencies to perform the cognitive activities associated with critical thinking. This is not empty. It implies that someone who purposefully behaves like a critical thinker may not actually be a critical thinker *even while behaving like one*, because she does so very rarely. Other definitions, however, suggest a stronger interpretation: Critical thinking requires independently definable traits or persisting individual differences, e.g., in variables like open-mindedness. Such variables might then affect one's ability to learn or consistently use critical thinking strategies:

Author	**Critical thinking involves specific dispositions and attitudes**
Dewey (1910/1991)	* Reflective thinking… involves overcoming the inertia that inclines one to accept suggestions at their face value…the most important factor in the training of good mental habits consists in acquiring the attitude of suspended conclusion…
Reeder (cited in Govier, 1987, pp. 238-9)	*… This focusing is not adventitious, but results from a conscious decision to think about or think through the things one encounters, and to develop habits which promote the implementation of such a decision.
Glaser (1941; quoted in Walters, 1994)	An attitude of being disposed to consider in a thoughtful way the problems and subjects that come within the range of one's experience…"
Facione/American Philosophical Association (1990)	* 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.
Paul (1993)	A unique kind of purposeful thinking in which the thinker systematically and habitually imposes criteria and intellectual standards upon the thinking… (p.21)
Wade & Tavris (1993)	The ability and willingness to assess claims and make objective judgments on the basis of well-supported reasons.
McPeck (1994))	The skill or propensity to engage in an activity with reflective skepticism
Siegel (1997)	Being a critical thinker requires basing one's beliefs and actions on reasons; it involves committing oneself to the dictates of rationality... (pp. 13-14) … Critical thinking has …a *critical spirit* component, which is understood as a complex of dispositions, attitudes, habits, of mind, and character traits. (p.27)
Browne & Keeley (1998)	* …3. desire to actively use the critical questions (p. 2)

What Are the Most Significant Variations in Current Usage?

A look at these definitions suggests that they do share, at a coarse level, an important common element: Critical thinking involves *the deliberate evaluation of intellectual products in terms of an appropriate standard of adequacy*. This is not so much a "definition" as a minimal core concept, a lowest common denominator. The specific definitions differ in how the statement is further fleshed out, but fleshed out it must be. These variations define a set of significant outstanding issues in the field.

Normative issues:

- What are the *intellectual products* that are evaluated? Must they be *beliefs* or can they include actions or the results of other intellectual activities? Must the *process* of thinking be evaluated as well as the *products* of thinking?
- What kind of normative *adequacy* is required? How is it determined? What are the appropriate *criteria* of evaluation? What is the role of *logic*? [12]
- To what extent does the application of the criteria vary with *context?* Are the criteria *general*, or are they relative to a particular *domain*? [13]

Cognitive process issues:

- Must the evaluation process include specific activities such as identifying the *implications* and *assumptions* embedded in ones' own views? Must one also confront the interests or biases that might affect one's judgment about a conclusion?

[12] If logical criteria are required, the only intellectual products to be evaluated are beliefs.

[13] Logical criteria are context-invariant and universal.

- Must the evaluation process include exposing one's own view to *challenge by others*? Must it include actively considering *alternative views*?
- Must one *actively seek out information*, or is it sufficient to judge based on the information already available?

Cognitive mechanism issues:

- To what extent must the evaluation process be *self-aware*? To what extent must the evaluation process be verbally *articulated* in the form of an argument? To what extent must critical thinkers draw on a *self-concept* of themselves as active shapers of their beliefs and thinking processes?
- To what extent is critical thinking *effortful* or associated with negative affect?
- To what extent must critical thinkers have appropriate *traits* or *attitudes*, e.g., persistent critical thinking behavior, or dispositions to suspend belief, be open-minded, or adopt a skeptical attitude?

Among the issues we raised earlier, in a spirit of devil's advocacy, was whether critical thinking in an Army context would take too much time or reduce confidence, and whether it might dilute the benefits of experience or stifle innovation. If rigidly applied, some variants of critical thinking outlined in this chapter threaten to have just such consequences. For example, must reasons for and against a point of view be explicitly considered for every belief and every action? Must critical thinkers always reflect explicitly on the criteria used to evaluate such reasons? Must the evaluation of reasons include consideration and rejection of every implicit assumption? Must alternative viewpoints be explicitly considered on every occasion?

Clearly, a reasonable concept of critical thinking should be useable. And to be useable, it must provide some flexibility in the way actual decision makers navigate through such options. But no current model shows how this can be done.

5. STRATEGIC CHOICES: INSIDE VERSUS OUTSIDE

In the previous chapter, we noticed significant variations in current definitions of critical thinking. In this chapter, we will address some historical and philosophical assumptions that can be discerned beneath these differences. *Why* have certain choices among the definitions of critical thinking been made rather than others? What specific combinations of normative and/or cognitive features fit together coherently and which do not? What are the deeper assumptions that glue together the components of alternative conceptions of critical thinking? What combination of assumptions has the most promise? To address these issues, it will be necessary to go beyond consensual definitions or *ad hoc* lists of skills to be trained.

We will find that a crucial geological divide in the field of critical thinking corresponds to two different views of rationality. On the mainstream view (*thinking from the inside*), rationality is reasoning according to correct rules, e.g., of logic, mathematics, and decision theory, that apply to explicit internal representations. On the other view (*thinking from the outside*), rationality is successful adaptation to the environment – regardless of how the adaptation is achieved and regardless of the cognizer's explicit thoughts about it. These two points of view, which are known in epistemology as *internalism* and *externalism*, respectively, underlie the most dramatic differences in assumptions about critical thinking.

The points of view are distinguishable at the normative level, by different assumptions about the *purpose* of critical thinking, the *constraints* under which that purpose must be achieved, and the *functions* that are necessary to achieve the purpose under those constraints. Moreover, these normative differences strongly influence theorizing about cognitive processes and cognitive mechanisms. As a result of this influence, variations among definitions at the cognitive levels can be largely accounted for by differences in normative assumptions. In turn, normative and cognitive variations account for differences in training strategies. Normative assumptions thus turn out to be the *building blocks* for putting together a critical thinking theory and training strategy. Choices at the normative level will have a major impact on the utility of critical thinking for the Army and elsewhere.

Are Criteria Needed for Critical Thinking?

There is, as we have seen, a rough consensus that whatever else critical thinking may be, one of its necessary functions is to evaluate intellectual products in terms of normative adequacy. That consensus will be our starting point:

Normative definition of critical thinking #1. Consensual

Purpose	*Accept an intellectual product if it is normatively adequate; reject it if it is inadequate.*
Constraints	*[Something here that narrows down possible ways to accomplish the purpose]*
Function	*Evaluate the intellectual product in terms of its normative adequacy.*
	Accept or reject the intellectual product based on that evaluation.

From this rather bland beginning, many thinkers move on to a stronger claim, that critical thinking involves the application of evaluative *criteria*. For example, if the intellectual product

under consideration is a belief, then criteria of normative adequacy specify properties of the belief, its evidential relationships to experience or to other beliefs, and perhaps additional facts about the situation, all of which combine to rationally justify that belief (Hunter, 1992, p. 82). Application of criteria is a good candidate for a normatively *necessary function* of critical thinking, because it is necessary from both a theoretical and applied point of view.[14]

The need for criteria is based on the requirement that an evaluative judgment have implications beyond a single case. If critical thinking is the process of rationally discriminating good from bad intellectual products, there must be *some* feature of the product and/or of the situation that can be appealed to as the basis for the discrimination; the label "good" or "bad" itself cannot be the only thing different. If it were, there would be no way to rationally contest or defend claims about the normative adequacy of beliefs. Two critical thinkers could rationally accept different evaluative conclusions about the same belief even though they agreed about all the other facts, indeed, even though they had precisely the same experiences and beliefs in all other respects. In this case, notice also that *critical discussion* is fruitless. No matter how diligently they probe, the two individuals will never discover a *reason* for the disagreement between them. Since they already have the same beliefs but one, there is nothing either can appeal to in order to persuade the other regarding the belief on which they disagree. Normative adequacy with respect to that belief would be in the eye of the beholder (Siegel, 1997: p. 20). There is an exact parallel between individual processes of critical thinking and social processes of persuasion and critical discussion. To defend one's beliefs against challenges from others, as well as to think critically about one's own beliefs, a person must apply criteria of normative adequacy that can generalize to new cases. If there are no such criteria, there is nothing to argue or think critically about.[15]

Ambiguity Versus Simplicity

Criteria, then, are facts about an intellectual product that provide as unambiguous a means as possible for judging its normative adequacy. To avoid ambiguity, i.e., room for disagreement among different assessors, it is necessary to specify identifiable features that discriminate good from bad intellectual processes. An additional requirement, for the criteria to be useable, is that they must not be excessively complex. Unfortunately, however, simplicity and lack of ambiguity are hard to achieve at the same time; one is usually purchased at the expense of the other. For example, it does not help much to say that the criterion of adequacy for a belief is that it be *justified*. Even though this is quite simple, no features are identified that could be used to anchor the normative concept of justification in facts. At the other extreme, it might be possible in principle to give a very detailed physical description (down to the last molecule if necessary) of a specific case of a justified belief. Such a description is unambiguous but maximally complex and non-generalizable. In looking for intermediate ground, it helps a little to break justification down, as informal logicians do, into *acceptability* of premises and *relevance* and *sufficiency* of evidence. In place of sufficiency, it might be a little more helpful to say, *I*

---

[14] Johnson (1996), states that, "if critical thinking does involve…evaluation of an intellectual product, then reference to criteria will necessarily be involved."

[15] At the very least, this requires that if precisely the same situation were somehow to recur, the evaluative judgment would have to be the same (Sosa, 1991, p. 110). A technical term for this is that normative judgments *supervene* on physical differences. The description of the situation, however, might be very complex and thus not practicably useable as a criterion.

can't imagine the premises' being true and the conclusion's not being true. Another way of breaking down sufficiency might be: *This is the only reasonable alternative of those that I have considered given the evidence that Ii have. I can't think of any other possibilities*. Such criteria are reasonably simple. There is room for ambiguity because they depend on evaluative terms like *acceptable, sufficient, relevant, reasonable,* and *evidence* and leave open the precise point at which one concludes that it not possible to imagine additional possibilities.[16] Nevertheless, they are informative and reduce ambiguity, by narrowing down the range of features that count.[17]

A small subset of beliefs appear to be susceptible to criteria that are *both* simple *and* relatively unambiguous (although even here it is not easy to state the criteria completely and exactly). This class of beliefs is supposedly acquired by *reflection* on the contents of one's own mind:

1. *Beliefs based on logical or mathematical intuition tend to be justified (e.g., 1 + 1 = 2).*
2. *Introspective beliefs about our own current thoughts and memories, visual experiences, and other sensations (e.g., pain) tend to be justified.*
3. *Beliefs deduced by logical rules (see point 1) from justified premises (see point 2) tend to be justified.*

Because of the simplicity and precision of such criteria, the traditional framework for critical thinking was based on reflective beliefs of these kinds. Unfortunately, such beliefs are only a small subset of all our beliefs. Not included are beliefs about the table in front of me, the table I can't see in the next room, the location of artillery, previous enemy actions, enemy intent for the future, and even my own identity and past history. Efforts to show how these beliefs might be supported by reflection have generally come to a dismal end, leading directly to skepticism. Internal reflection is surely important in critical thinking, but reflection alone doesn't help much in evaluating most instances of justified belief. A major reason for exploring the idea of critical thinking as *dialogue*, which we will pursue later, is the possibility it affords of formulating a more generally applicable set of criteria, which are nonetheless relatively simple and unambiguous (cf., Walton, 1998, pp. 7-9).

Criteria are, ideally, unambiguous and cognitively ascertainable specifications of the features that make beliefs *normatively* adequate, and which may be used to provide feedback in *training* (e.g., Sosa, 1991: pp. 178-181). Thus, criteria are links between normative, cognitive,

---

[16] Two potential misunderstandings should be addressed. First, requiring unambiguous criteria of normative adequacy, which contain no vague "evaluative" terms, is not the same as *reducing* normative adequacy to physical or descriptive differences. *Why* certain criteria and not others are correlated with normative adequacy is a normative rather than a physical / factual question. Second, this requirement does not fall prey to the "naturalistic fallacy," i.e., confounding what *is* with what *ought* to be. It does not imply that the correct unambiguous specifications describe actual behavior. (In *naturalistic* approaches, by contrast, the way people actually think *is* part of the rationale for a normative specification. But that is separate from the requirement that the specification be unambiguous.)

[17] Here's an example of how criteria can reduce ambiguity by narrowing down the range of properties to focus on: *He's the best basketball player in the league.* There are some facts that everyone would acknowledge are relevant (points scored; assists; free throw percentage; defensive points), but not everyone would identify exactly the same facts or give them the same importance; moreover, some of the relevant facts might themselves be somewhat subjective (e.g., does he show leadership on the court?). We therefore do not have a rule that guarantees the same judgment from every evaluator. Nevertheless, a considerable amount of guidance is available, so the results should not be totally unpredictable either. Is this as good as it gets with respect to criteria for evaluating beliefs?

and applied levels of analysis. Two general constraints (at least) must be satisfied by criteria: First, if criteria are used to evaluate intellectual products, they must be usable by the critical thinker and/or evaluator; hence, they must be accessible to consciousness and sufficiently simple and unambiguous to use. Second, satisfaction of the criteria must be correlated with normative adequacy, that is, the criteria tend to fit an intellectual product if and only if that product is normatively adequate. We modify the consensual definition accordingly:

Normative definition of critical thinking #2. Making the role of criteria explicit

Purpose	*Accept an intellectual product if it is normatively adequate; reject it if it is inadequate.*
Constraints	*(1) There is an association between certain facts about an intellectual product and its normative adequacy.*
	(2) These facts are ascertainable by the evaluator.
Function	*(1) Determine which of the relevant facts are true of the intellectual product.*
	(2) Evaluate the normative adequacy of the product in terms of those facts.
	(3) Accept or reject beliefs as a function of that evaluation.

Theories of critical thinking differ on what facts should be used as criteria and on the points of view we should use to ascertain them.

What Is the Assessor's Point of View?

A second, quite different purpose for criteria is the evaluation of thinking in a training or educational context. We must know how to distinguish good from bad critical thinking in order to train critical thinking skill. Indeed, unless we can assess critical thinking against some standard, it cannot be regarded as a cognitive skill at all (Lipman (1987). It is important, however, to distinguish the use of criteria by critical thinkers, considered in the previous section, and the use of criteria by critical thinking *trainers*. Trainers of critical thinking have a different point of view from those they evaluate, including differences in objectives, available information, and time constraints. A critical thinking trainee (or practitioner) is directly interested in whether or not to accept a claim, but the trainer is not. The trainer is directly interested in evaluating the trainee's thinking performance and/or in providing corrective feedback for its improvement. The educator uses criteria for evaluation *of* thinking, while the critical thinker / trainee uses criteria for evaluation *in* thinking. As a result, the criteria themselves may be different.

The application of criteria by a critical thinker can be thought of as part of the interplay between a *first person* and *second person* point of view in a simple persuasion dialogue (e.g., van Eemeren & Grootendorst, 1992; Walton & Krabbe, 1995). For present purposes, the relevant rules are simple

1. The first person proposes a conclusion,
2. The second person either challenges the conclusion or concedes it,
3. The first person either defends the conclusion (e.g., with a reason) or retracts it,

and so on. In this kind of interchange, both the first and second person (the proponent and opponent, respectively) must evaluate the credibility of the conclusion at each turn in the dialogue. We have already discussed some criteria of this kind, e.g.:

> *Accept (or do not challenge) a belief when other beliefs you accept are sufficient to show that it is true.*

A less ambiguous example was this:

> *Accept (or do not challenge) a belief when it is the most reasonable alternative that you have thought of given your other beliefs, and you can't think of any additional, competing possibilities.*

Here is a much less ambiguous example of a criterion from the first or second person point of view:

> *Accept (or do not challenge) the belief that there is a fire when you see smoke.*

A problem with this criterion, of course, is that it is not very general (even if it were always true). Simplicity is lost because too many such criteria would be necessary. Pattern recognition processes may in fact operate, in part, with innumerable highly specific learned patterns of this kind.

Evaluations by the proponent and opponent are *internal* to the dialogue process. The trainer's point of view is quite different, and so are her criteria. She acts as the observer of a belief generation process, and assesses it from a *third person*, external perspective. For example, the trainer / evaluator might apply criteria such as the following:

> *Beliefs tend to be correct when they result from a dialogue process structured by the following rules:…*
>
> *Beliefs tend to be correct when they result from a reasoning process structured by the following rules:…*
>
> *Beliefs tend to be correct when based on recognition by a highly experienced decision maker in a situation familiar to her.*[18]

In short, criteria are always applied from a *point of view*. In presenting, challenging, and defending her own beliefs, the critical thinker alternates between first and second person points of view, while the training evaluator adopts a third person perspective on that interaction in order to improve the process as a whole. The different roles played by first, second, and third person points of view are illustrated in Figure 5.

Just as the critical thinker can adopt the second person stance toward herself in order to challenge her own reasons and conclusions, so she can adopt a third person stance toward herself in order to evaluate the process she is using. The critical thinker's purposes are similar if not identical to those of the trainer – to identify the type of thinking process that is appropriate for a particular situation, to monitor its reliability and efficiency, and to learn ways of improving it.

---

[18] This illustrates that even third person criteria can vary in ambiguity. This criterion falls short unless we are better able to specify the requisite degree and type of experience, and the relevant metrics of similarity among familiar situations in different domains.

Thus, the third person point of view is useful not just to the trainer, but to the critical thinker herself. Critical thinking emerges out of the interplay among all three of these perspectives.[19]

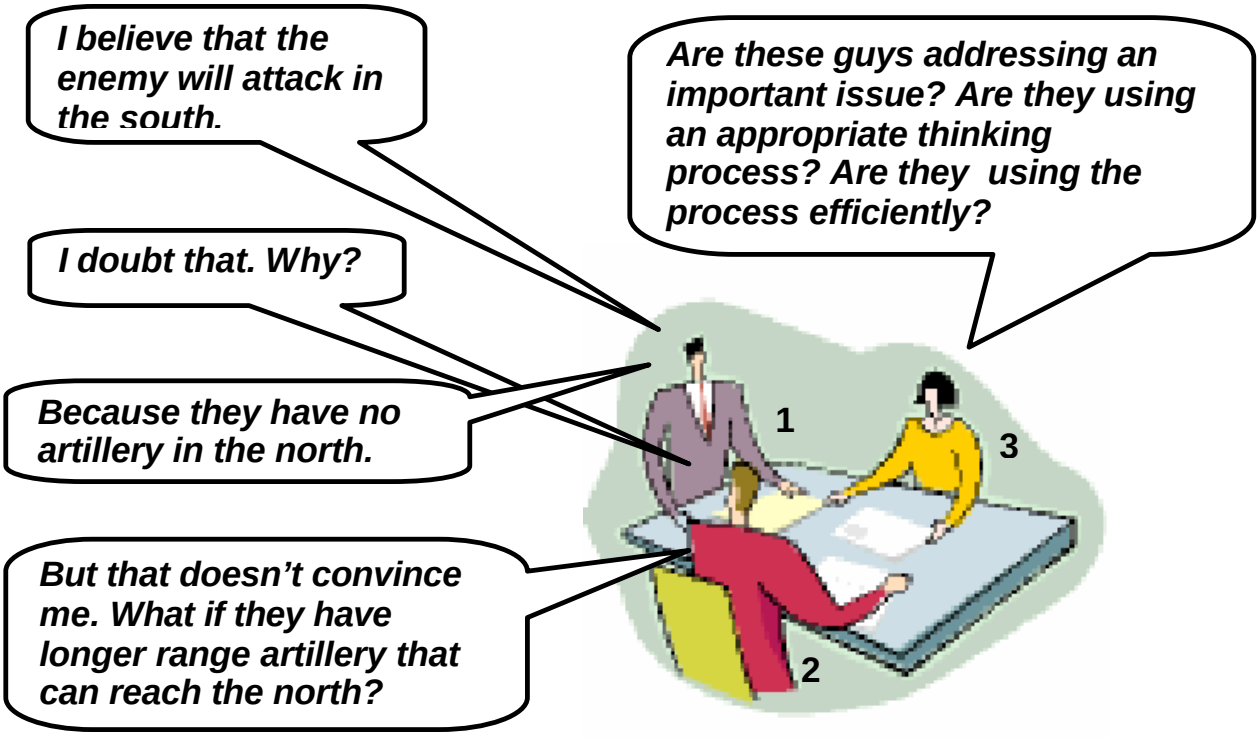


Figure 5. Three points of view in critical discussion: First person, second person, and third person.

In addition, what if the critical thinker needs to evaluate beliefs that are not based on reasoning but on perception or recognition? The usual interplay between first and second person will be stymied, because the thinker may be unable to articulate explicit reasons for accepting the belief or decision. In such cases, if critical thinking is possible at all, it *must* be "from the outside," e.g., based on an assessment of the reliability under the prevailing conditions of the perceptual or recognitional processes that led to the belief in question (Goldman, 1979). Even when conclusions are based on inference, many beliefs that play a role in the reasoning remain implicit. The cognizer must fall back on an assumption about the reliability of the processes

---

[19] Here is a version of the dialogue (Figure 5) in which a single cognizer shifts among the three perspectives:

First person role proponent): *I believe that the enemy will attack in the north*.

Second person role (opponent): *But why do I believe that?*

First person role: *Mostly because the enemy have no artillery positioned in the south.*

Second person role: *On the other hand, I can't rule out the possibility that they longer range artillery.*

Third person role (judge): *Is this important for me to think about right now? Am I using an appropriate thinking process? Am I using the process efficiently?*

underlying the mutual interaction of beliefs. An external or third-person view of oneself as a cognizer appears to be a key component of critical thinking.

Where Did the Idea of Critical Thinking Come From?

One of the most prominent themes of early modern philosophers such as Descartes, Locke, Berkeley, and Hume, was the importance of challenging inherited and customary beliefs, that is, to adopt not only a first-person but also a second-person (critical) point of view. This imperative to *doubt* one's own accepted beliefs gave birth to the idea of critical thinking, and this view of critical thinking gave birth to early modern philosophy. The pragmatist movement (e.g., Dewey) introduced a third-person perspective, which defined rationality instrumentally as successful achievement of goals. But the influence of its earlier origins remains very strong. If we ignore these roots, we will miss their persisting influence, and we may end up seeing critical thinking through the filter of 17th century ideas.

The theory of knowledge (*epistemology*) was a key concern of Descartes, Locke, Berkeley, and Hume. They agreed on two things about thinking:

1. Its purpose is to fulfill an ethical duty to *think properly* about whether to accept or reject each of our beliefs (Plantinga, 1993a, pp. 3-29)
2. A constraint on proper thinking about belief acceptance is that it must be based upon good *evidence*.[20]

Initially, evidence was regarded as sufficient only if it guaranteed the truth of a conclusion. Today, philosophers and critical thinking theorists acknowledge uncertainty about matters of fact and even about logic. The purpose of critical thinking is now seen as more modest, to ensure a high *probability* of truth. In retrospect, however, the introduction of uncertainty is less important than what has survived intact from this tradition: Critical thinking is still thought of as akin to inner ethical goodness, based on doing one's duty as judged by an inner light rather than by the external consequences of one's actions. To accept beliefs only when they are seen on reflection to be the conclusions of good arguments is correct mental hygiene, regardless of conditions or outcomes in the real world. The lingering effects of this *internalist* tradition are responsible for many of the features (and shortcomings) of current theories of critical thinking. As we shall see, it limits the permissible interplay between first and second person voices in critical dialogue, and eliminates the possibility of a third-person perspective altogether.

Another influential philosophical tradition besides epistemology was *formal logic*. According to this tradition, a *necessary function* of critical thinking is the application of logical criteria to assess the relationship between reasons and conclusions. Just as contemporary

---

[20] Compare the following statements by Descartes and Locke with the contemporary definitions of critical thinking we looked at earlier:

> … if I abstain from giving my judgment on any thing when I do not perceive it with sufficient clearness and distinctness, it is plain that I act rightly… But if I determine to deny or affirm…even though I judge according to truth, this comes about only by chance, and I do not escape the blame of misusing my freedom… (Descartes; quoted in Plantinga (1993a, p. 12).

> … a firm assent of the mind… if it be regulated, as is our duty, cannot be afforded to anything, but upon good reason… He that does not this to the best of his power, however he sometimes lights on truth, is in the right but by chance… (Locke; quoted in Plantinga (1993a, p. 13).

epistemologists have acknowledged uncertainty, so-called *informal logicians* have moved away from formal logic, i.e., criteria based solely on "form" or syntax. But the logical tradition also has had a persisting influence: This is reflected in the notion, held by informal logicians and other critical thinking theorists, that reasoning transpires by means of *arguments* that use current beliefs as premises to justify the acceptance of new beliefs as conclusions. Arguments are evaluated by making implicit premises explicit and then applying explicit criteria to the relationship between the premises and the conclusion. As we shall see, there is good reason to question the universal applicability of such an argument strategy. In interesting cases of uncertainty, cognizers must evaluate sets of beliefs as a package, in terms of the coherence of the whole, rather than moving linearly from already accepted beliefs to new ones. Evaluation of explicit beliefs, moreover, always draws on a vast store of implicit background knowledge. Thus, the "argument" can never be made fully explicit. Ultimately, dialogue models may provide a deeper and more comprehensive way to classify the different types of reasoning that people in fact engage in. But to appreciate this in full, we first need to examine the traditional assumptions.

What Assumptions Does the Traditional View Make?

The effects of traditional epistemological and logical assumptions on the critical thinking movement have been profound. Once we accept the purpose and constraint – that we have a quasi-ethical duty to base beliefs on evidence – the rest of the traditional paradigm for critical thinking follows very quickly. This paradigm – which has held strong from Descartes to the present day – views critical thinking as the responsibility of a single individual who is reflecting self-consciously and logically on her own thoughts.

Internalist Constraints on Evidence

It is unfair to hold someone responsible for the performance of a duty that is not within her power to fulfill. So, if it is our duty to accept only justified beliefs, then we must be able to know whether or not a belief is justified for us (Plantinga, 1993a; Steup, 2001) and to accept it or reject it accordingly. If justification depended on information to which we might or might not gain access, or on the collaboration of other individuals, success or failure would be attributable to good or bad luck. Therefore, whether or not a belief is justified for a person depends only on evidence to which *that person* has direct conscious access. All the factors upon which a person must rely to determine the justification of a belief must be available to her by *reflection* on the present contents of her mind (the top right cell in Figure 4). She can form beliefs about them simply by attending to the contents of her awareness. A belief that fits all the evidence available by reflection to a person at a given time can never turn out *not* to have been justified at that time for that person (even though, if we accept uncertainty, it might, turn out to be *false*). This position is called *internalism.* According to one of its leading proponents (Chisholm, 2000, p. 119):

> The internalist assumes that, merely by reflecting upon his own conscious state, he can formulate a set of epistemic principles that will enable him to find out, with respect to any possible belief he has, whether he is justified in having that belief.

But why is justification by evidence necessary or desirable in the first place? To answer this, internalists try hard to establish a link between justification based on *internal* evidence and *external* truth, or knowledge. As we saw earlier, to be fair in holding someone responsible for the

justification of her beliefs, it must be within her power to hold only justified beliefs. That is why evidence must be internal. By the same token, to hold her responsible for *knowledge*, it must be within her power to accept only beliefs that are true (or, in more recent times, likely to be true). If a belief is true (or likely to be true) purely *by accident*, with no particular reason to have expected it to turn out true, it is not knowledge. (A correct prediction of the weather based on sheer guessing is not knowledge.) In other words, if a true belief is to count as knowledge, the cognizer must have *good reason for expecting the belief to be true*. This is to say that the cognizer must have evidence that *increases the probability that the belief is true*. But *justification* can be defined as evidence that increases the probability of a conclusion (P. Klein, 2000). Defined in this way, justification is necessary for knowledge as long as it must be fair to hold someone responsible for their knowledge.[21]

The traditional internalist paradigm is individualistic: The only way a belief can be justified is for an individual who owns that belief to reflect on the contents of her own mind. If two individuals happen to be aware of different evidence, one may be justified in accepting a particular conclusion while the other is not. Does the internal character of evidence mean that justification is subjective, that whatever any person believes to be justified *is* justified for that person? Some recent critical thinking theorists have in fact adopted relativist perspectives on knowledge. But the internalist tradition resists this. To ensure the link between justification and truth, internalists combine individualism with universality. They insist that the features of internal evidence that qualify it as evidence and the relationships between evidence and conclusion that constitute justification, are the same for everyone, independent of context. According to Feldman and Conee (1985/2000) and Siegel (1997), justification consists in the *objective fit* between consciously available evidence and a conclusion. Since evidence makes a conclusion more probable in some objective sense, any individuals *with the same evidence* are justified in accepting the same conclusions.

Here is an internalist normative definition of critical thinking that brings together these themes:

[21] Probability in the relevant sense is not merely a subjective degree of belief, as in Bayesian theory. It must correspond to objective facts, e.g., to actual frequencies or propensities (Pollock & Cruz, 1999).

Normative Definition of Critical Thinking #3. Internalist

Purpose	*To accept beliefs that I am justified in accepting, and to reject beliefs I am not justified in accepting.*
Constraints	*(1) To justify a belief, it is necessary to show that it has a sufficiently high objective probability of being true based on consciously accessible evidence.*
	(2) Beliefs that fit the evidence have a greater objective probability of being true.
	(3) Facts about whether or not a belief fits the evidence are consciously accessible.
Functions	*Critical thinking is:*
	(1) the identification of consciously accessible facts about the evidence for beliefs,
	(2) context-independent application of criteria to determine how well the beliefs fit the evidence, and
	(3) acceptance or rejection of beliefs based on their fit to the evidence.

The principle challenge for internalism, clearly, is to specify the *criteria*. These criteria must apply to evidence that is *both* readily accessible within the cognizer's current conscious experience, *and* associated with higher objective probability of truth. Moreover, the internalist is expected to *understand* the impact of that internal evidence on the objective probability of the conclusion.

Logical Criteria

What kinds of relations do we look for when we evaluate the fit between some thoughts (the evidence) and other thoughts (the conclusion)? And what properties qualify some thoughts to be treated as evidence? What is this magical set of features that (i) can be identified by solo reflection, and (ii) should nevertheless be identified the same way by all individuals in all contexts, as an objective signals that a belief is (likely to be) true?

Beliefs can provide evidence for other beliefs only if they are occurrent, that is, if they are actively present in thought. Beliefs stored in long-term memory do not count as evidence because the cognizer may not see their connection to the current conclusion. A key feature of occurrent beliefs is that they have assertive propositional content. They stake a claim that something is the case, and thus they may be true or false. Visual experiences (e.g., the sensory appearance of a tank) count as evidence only if they assert some propositional content (e.g., *there is a tank*) and thus are, in effect, visual beliefs.[22]

[22] A strong line of internalist opinion insists that if inner events are not propositional, they cannot serve as evidence. A visual experience, sensory memory, touch sensation, imagined experience, or a pain can no more support a conclusion than any other non-propositional object. In a courtroom, a physical object, such as a gun, may be referred to as "evidence." Strictly speaking, however, it is not the gun per se that supports a conclusion or justifies a verdict, but propositions about it: e.g., beliefs that the jury forms about the gun based on their perception of it and their evaluation of statements (e.g., by police that this gun was found in a certain place). It is in deciding whether to

To determine whether a belief is justified, a critical thinker must examine consciously accessible *facts* pertaining to that belief. Two kinds of facts are potentially relevant:

- *Intrinsic properties* of the belief, e.g., the fact that it is a vivid visual belief or that it is a logical truth.
- *Relations* the belief has to other beliefs, e.g., the fact that it logically follows from other occurrent beliefs.

Internalist theories differ on the scope and importance of *intrinsic properties*, but all theories recognize the importance of *inferential relationships* among beliefs.[23] They also tend to agree that criteria for evaluating inferential relationships must satisfy internalist constraints (as shown in the normative definition above). Internalist constraints apply both to the *backing* (or source) of the criteria and the *grounds* for their application.

The first issue is the *backing* or source of the criteria: how we discover the right criteria to evaluate inferential relationships. Suppose that the epistemic principles that identify criteria were based on empirical observation or scientific findings. For example, in order to know what counts as a correct inference relation between evidence and conclusion, we had to have certain kinds of learning experiences, or be aware of the success or failure of theories or research strategies in a particular domain. If so, a decision maker would be out of luck if she did not happen to have the relevant knowledge. But internalists demand that justification be within the power of the cognizer and not a matter of luck. According to them, it would not be fair to hold someone responsible for accepting unjustified beliefs or rejecting justified ones if she had no access to the necessary information. To be fair, we must insist on criteria whose relevance is known or readily knowable to all cognizers.

The second issue pertains to the *grounds* for asserting that the criteria are or are not satisfied in a particular case. The grounds cannot go beyond the information in our current awareness about the relationships among occurrent beliefs. The relevant relationships must be identifiable from reflecting on the propositional contents of the beliefs themselves, without reference to other facts about the situation or cognizer.

---

accept or reject these propositions that the jury determines the evidentiary weight of the gun. If such propositions are *true*, they may be used to narrow down the range of plausible possibilities regarding the truth of other propositions (e.g., about who committed the crime). Ideally, the only surviving possibilities will be those in which a relevant conclusion (guilt or innocence of the defendant) is the case. But the gun *per se* can be neither true nor false, and does not directly narrow down the range of possibilities. Both evidence and conclusion must be beliefs capable of truth and falsity.

But can't non-propositional sensory experiences "support" beliefs in the sense of *causing* them to be held? Isn't this a way of transmitting information from the world to the brain? Such causal links do not belong in an internalist account since they are not directly accessible to consciousness. Moreover, according to internalists, even if causal connections could be internally "observed," the existence of a causal connection is not sufficient to show that a perceptual experience *justifies* a belief. For that, some kind of logical relationship is required. Internalists generally reject naturalistic approaches to justification or knowledge, because they invoke cognitive mechanisms that are both external to consciousness and non-normative. Fumerton (2001) and Bonjour (2001) try to show that perceptual experiences may be basic without being beliefs, but for both Fumerton and Bonjour, the perceptual experiences must have content of which we are aware, hence, are belief-like.

[23] As we shall see, there are two main schools of internalists. For *foundationalists*, both (a) intrinsic properties and (b) relations to other beliefs count toward justification. *Coherentists* recognize only (b) relations to other beliefs.

Taken together, these constraints on grounds and backing point to the unique importance of *formal logic* in internalist theory. To see how the constraints work, consider the following inference:

1. *The enemy has not placed artillery in the north.*

2. *If the enemy has not placed artillery in the north, they will not attack in the north.*

3. *Therefore, the enemy will not attack in the north.*

First, the internalist paradigm demands criteria with non-empirical backing. That is, they transmit justification from evidence to conclusion in some *necessary* way, not dependent on empirical facts about the world that can only be learned through observation and/or theory. The simple inference above satisfies this constraint. It fits a general inference schema called *modus ponens*, which is valid regardless of the specific content of the propositions involved. Logical rules are thought by some philosophers to be knowable *a priori*, without need of backing in experience. No matter what sentences are substituted for *p* and *q* in the pattern below, the reasoning is valid:

1. *p*

2. *If p then q*

3. *Therefore, q.*

Second, the internalist paradigm demands criteria that can be applied to internally accessible information about the relevant propositions (i.e., grounds). The inference above also satisfies this constraint, since the *modus ponens* schema is applied to the form or syntax of propositions, independent of their specific meaning, referents, or context. Criteria based on formal logic thus appear to uniquely satisfy both internalist constraints.[24]

Reflective Mechanisms

The internalist paradigm emphasizes voluntary choice of beliefs based on reflection. This exercise of choice may not be particularly easy. It may require substantial effort to root out long-held or popular beliefs, and it may sometimes be extremely difficult to resist jumping prematurely to conclusions, e.g., to reject a compelling perceptual or cognitive illusion. But internalists assume that it is within the power of the cognizer to do so. Unless acceptance and rejection of beliefs is voluntary, we cannot hold someone responsible for her beliefs. The internalist position thus has implications for the role of cognitive mechanisms, particularly, the role of consciousness, choice, and mental effort. The combination of those three elements approximates a faculty that philosophers used to call the "will." Critical thinking is an exercise of the will; it resides in the reflective, deliberate, effortful dimension of thinking.

---

[24] In recent years, both of these traditional claims about formal logic have been strongly disputed. In terms of backing, Quine and Ullian (1970) and Everitt & Fisher (1995) argue that logic is part of our overall theory of the world and is thus not known *a priori*. Logical principles, like empirical theories, might be revised under pressure from observational and experimental findings (e.g., in quantum physics). In terms of grounds for application, an enormous amount of judgment is required to parse real-world statements in terms of their logical form. Woods (2000) shows that the "logical form" of a particular proposition depends on the logical theory that we choose to apply to it. Thus, the traditional internalist appeal to logic is mistaken. There are *no* criteria that satisfy the internalist constraints.

A major bone of contention in internalist theory is *how much* reflective awareness is necessary for justification. Unfortunately, from the internalist point of view, having internally accessible thoughts that satisfy appropriate criteria turns out to be insufficient for justification. For example, in the simple illustrative inference above, suppose all three propositions are occurrent beliefs. Clearly, the relation between them is logically tight; the third proposition follows by a valid inference from the first and second. We might think that this is enough to show that belief 3 (q) is justified if beliefs 1 (p) and 2 (If p then q) are known to be true. But suppose the cognizer does not *see* the logical connection, and happens to accept proposition 3 for reasons altogether unrelated to 1 and 2 (e.g., a psychic told her that 3 was true). Unless she understands *why* the evidence supports the conclusion, the conclusion is not justified for her. She might be thinking simultaneously about the evidence and the conclusion by coincidence and not see any connection at all. Or she might misunderstand the connection between the evidence and the conclusion, i.e., by using an invalid logical rule that happens to give the right answer in this instance. In such cases, even though the cognizer has occurrent beliefs for both the evidence and the conclusion, and the evidence stands in the right relationship to the conclusion, her belief in the conclusion would be correct only by accident (Bonjour, 1985; *pace* van Cleve, 2000, and Allston, 2000), and thus is not justified.

A basic principle of internalism is that justification is completely determined by internally accessible evidence. Thus, if one individual is justified by her internal evidence in holding a belief while another individual is not, they must have different internal evidence. The inevitable internalist tactic, then, is to look for flaws in the evidence for proposition 3 in the above inference. Simply believing 1 and 2 alone is clearly insufficient, since it does not distinguish someone who correctly believes 3 *on the basis of* 1 and 2 from someone who does not. Thus, internalists sometimes insist that reflection at a *higher level* is necessary before a belief is accepted. Not only must the evidence be adequate, the cognizer must be explicitly *aware* that the evidence satisfies the relevant criteria. The illustrative inference must now be expanded (using letters in place of sentences for brevity):

1. *p*

2. *If p then q*

R. *If p and (If p then q) then q.*

3. *Therefore, q.*

where the new premise, R, makes explicit the inferential relationship between the original two premises (1 and 2) and the conclusion (3). Unless the cognizer is aware of this relationship, as formulated in R, the conclusion is not justified for her. The requirement for self-conscious reflection is striking: On some internalist views, no first-order belief is justified unless there is a *meta-belief* that the evidence is sufficient for the conclusion with respect to appropriate criteria (Sosa, 1991: pp. 181-183).[25]

[25] The obvious problem with this proposal is that it generates an infinite regress. After all, one might have the appropriate first- and second-order beliefs on the list, but not see the connection between *them* and the conclusion. A third-order beliefs would be necessary, i.e., R′ *If p and (If p then q) and (If p and (If p then q) then q) then q.* But the same problem arises again, showing that a fourth order belief R″ is necessary, and so on. (See footnote 37.)

The traditional paradigm combines looking inward for evidence, reliance on strict logical criteria for inferring new conclusions from the evidence, and a demand for self-conscious reflection and voluntary control over belief at every step of reasoning. From the traditional point of view, rational thought, correctly carried out, *requires* critical thinking. Critical thinking is precisely the attitude that Descartes and Locke promoted: the appropriate exercise of the will to withhold belief unless we are consciously aware of evidence that is logically adequate (and is *seen* to be logically adequate) for the conclusion. We may fairly refer to this combination as the "intellectualist model of justification" (Sosa, 1991, p. 195; Pollock & Cruz, 1999). But is this the most viable account of what critical thinking is all about?

What Does the Internalist Paradigm Look Like?

From the initial goal of challenging superstition and dogma, modern philosophers have spun a complex web. Perhaps no single author subscribes to all of the assumptions they adopted, but they still exert a strong gravitational attraction on theoretical choices, as evidenced by the definitions in the critical thinking literature that we surveyed earlier. Here are the answers given by the intellectualist paradigm to the questions about critical thinking that were left open by our consensual definition: [26]

Normative issues

- Does critical thinking apply to intellectual products other than beliefs? The concept of *duty* in fact originates in the realm of *action* rather than belief. It applies to ethical standards of right action based on inner *intent* (e.g., Kant's categorical imperative) rather than on actual or expected consequences (e.g., utilitarianism). Thus, internalist criteria apply most naturally to ethical action in this Kantian sense. They have been extended to beliefs, but only at the cost of an assumption: that internally accessible evidence exists which is necessarily linked to external success, i.e., justification and ultimately, truth. Internalist criteria have not been extended to *instrumental* actions, which are undertaken as means to desired ends, or to other intellectual products, such as works of art or stories. The success of such actions and products is determined by external causal relationships between actions and consequences.
- Does critical thinking evaluate thinking processes, or only their products? Cognitive processes extend over time, while internalist justification is based on a snapshot of the relationships among beliefs in momentary awareness. Past events, including past thoughts, do not count toward justification because successful memory involves luck. The aim of the traditional paradigm was to eliminate the element of luck in identifying justified beliefs, and to eliminate (or reduce) the element of luck in hitting on true beliefs. The fact that I seem to recall previous steps in a process has the same status as other beliefs. The claim that such memories match the reality of what happened in the past will itself require justification, which is hard to come by. Thus, the occurrence of a temporally extended process cannot be essential to justification. In a process that may span a significant length of time, it is only the currently

[26] See the section above entitled, *What Are the Most Significant Variations in Current Usage?* Here are some of the issues: What does critical thinking evaluate? Using what standards of adequacy? With what degree of universality? What is the importance of reasons, alternative views, and active information seeking? What is the requirement, if any, for reflective awareness, effort, and enduring dispositions?

occurring thoughts that count. Critical thinking is thus confined to the upper right cell of Figure 4, i.e., the current contents of a specific, actually occurring process. Static properties and relations among thoughts are both necessary and sufficient to determine justification of a belief. This is at best an emaciated notion of *process* that discards both temporal *extension* and input-output *functions*, i.e., persisting disposition to act differently under different circumstances.

- Is critical thinking universal? The properties and relations used as criteria must be *general rather than domain-specific*. Otherwise knowledge of them would demand specialized expertise and people could be mistaken about what the criteria were or about how to apply them. Moreover, for the same reasons, the outcome of an evaluation must be the same for any individual in *any context* who has the same evidence. The criteria ignore any external information not known to the cognizer.
- The criteria prominently include *formal logic*. Logic applies only to the abstract, internal form of beliefs and (supposedly) applies universally and independent of context. Other, non-relational standards of acceptability may also be involved, such as supposed self-evident intuition, but they would also have to be universal and necessary.

Cognitive processes

- Cognitive processes are largely ignored by the intellectualist paradigm, because of the principle that chance should play no role in justification. This rules out *learned* strategies for achieving justification or finding relevant evidence (such as those depicted in the top row of Figure 4), since different individuals might have better luck at acquiring the relevant skills or in exercising them. The traditional paradigm insisted that all the relevant evidence must already be present in consciousness. Thus, it is unnecessary to use fallible strategies for directing attention, searching for relevant information to be retrieved from long term memory, or exploring the external environment. Active *information seeking* is irrelevant since all pertinent information must already be present among the cognizer's thoughts. Strategies designed to extract knowledge stored in long-term memory are irrelevant since long-term memory is a mechanism not directly accessible to consciousness.
- Critical thinking does not require the identification of *assumptions or interests*. Diagnosis of specific types or causes of mistakes is irrelevant. All that matters is whether criteria of correctness are or are not satisfied by one's current thoughts.
- Exposure to *challenge by others* is irrelevant. Justification of beliefs is the responsibility of the individual who has those beliefs. It is a lone enterprise, in which the cognizer reflects on her own thoughts. Confrontation with *other views* is irrelevant. If one's own beliefs and evidence satisfy the relevant criteria, there is nothing more to be learned by considering other views on the same issue. [27]

[27] Another consequence is that the process of challenging and defending views in dialogue with other individuals has no place as a reasoning paradigm. Access to such a dialogue must be through an individual's own beliefs, and they require internal justification which cannot itself involve dialogue on pain of circularity. Beliefs about the testimony of others can serve as evidence only if the cognizer has internally justified beliefs about the reliability of

Cognitive mechanisms

- Rationality requires higher-order *self-awareness* about the fit between arguments and criteria. We are treating self-awareness as a "mechanism" because that is the way consciousness is thought of in contemporary cognitive psychology. For the internalist, however, self-awareness was a conscious state internally accessible to the cognizer. The requirement for self-awareness was a primary normative requirement. Similarly for the next item, the expenditure of mental effort.
- Rationality often requires an *effortful* exercise of will to resist habitual or automatic beliefs. Again, we treat expenditure of mental effort as a mechanism, although for the internalist it would be a consciously experienced event or state.
- Critical thinking is not an optional mode of thinking but rather is obligatory if we wish to have justified beliefs. It is equivalent to rational thought. Thus, a strong *disposition* to think critically at all times is desirable. Nevertheless, the presence or absence of such a disposition does not enter into the justification of a particular belief. Justification of a belief depends in no way on enduring individual traits of the cognizer, only on the contents of her mind at a given time. There is no need for the cognizer to be a habitual *critical thinker* in order to be justified on a particular occasion.

In sum, according to the traditional internalist paradigm, critical thinking involves the application of universal *normative standards* through the use of (what we would regard as) *cognitive mechanisms* such as consciousness and mental effort. What is striking is that an entire level of analysis, *cognitive processes*, including adaptive cognitive strategies, is missing. The root cause of this neglect is the internalist insistence that justification involve ethical fairness, interpreted as the complete absence of chance, in assigning praise and blame.

Suppose we drop the idea that critical thinking is a matter of fairly assigning praise and blame based on duty as the cognizer sees it. If we do so, the way is opened for a broader, third-person view of critical thinking, which incorporates external factors that may not be known to the cognizer, which require the discovery and implementation of successful strategies for first and second person dialogue, and the exploration of the environment. We would have to acknowledge that an inevitable element of chance enters into the justification process and that it may not always be possible to guarantee a result. Strategies now move front and center: for information seeking, for retrieving and using more of one's own knowledge not currently in focal awareness, for identifying and mitigating specific types of errors or fallacies in reasoning, for collecting and analyzing data, for learning, and for critical debate.

What Is the View From the Outside?

Do you believe that in order to really know something, a person must always know *why* she thinks it is true? That she must always be able to give reasons and answer challenges? If so, you are an internalist, well within the intellectualist tradition that is the revered grandfather of critical thinking. Current views on critical thinking depart from it, if at all, in piecemeal and largely *ad hoc* fashion. A more effective and coherent paradigm for critical thinking may

---

the testifier. Reasoning used by another person is not binding on the cognizer unless she herself sees the logical relationships. Thus, there can be no *direct* justification-inducing property inherent in a dialogue process itself.

emerge, however, if we are willing to explore alternatives. A good place to look is the recent interest in *externalist* theories of knowledge.

Perhaps the most fundamental problem with internalism is that it sets too high a standard. Conscious reasons are not always necessary:

1. The intellectualist tradition has trouble explaining knowledge based on perception of physical objects or on memory for recent events. These are beliefs which may be mistaken (hence, are not self-evident) but for which we are ordinarily unable to point to reasons.
2. The internalist paradigm denies knowledge to experts who make what appear to be complex judgments relatively automatically. They may be so proficient in an area that they form judgments and recognize solutions to problems without being aware of reasons.
3. Internalism also fails to explain how beliefs may be justified even when reasons *can* be produced. Suppose that MAJ Nord arrives at a conclusion about enemy intent after thorough analysis. But the reasons for holding the belief are not at the moment in MAJ Nord's conscious mind. Many of them are in long-term memory, and some of them have probably been forgotten altogether (Harman, 1986). Moreover, the capacity of working memory might be insufficient to encompass them all at once (Goldman, 2001, p. 122). The internalists would regard her conclusion regarding enemy intent as unjustified! The fact that the supporting beliefs are in long-term memory is not good enough, because they cannot be ascertained by conscious reflection (except by actually retrieving them into consciousness).
4. Internalism also rejects the possibility of knowledge for animals and children. It would be impossible to say that a dog knows where its feeding bowl is, or that a child knows that her mother is present.

Ironically, because internalists view beliefs through the filter of duty, the vast majority of our beliefs turn out to be blameworthy! A paradox facing internalism is that (i) in order to assign praise and blame for beliefs fairly, we must evaluate them with respect to evidence that is in conscious awareness, but (ii) there is almost never enough evidence in conscious awareness to justify beliefs that should be acceptable. These include beliefs based on perception, memory, expert recognition, and information that is in long-term memory and/or partially forgotten.[28]

The internalist tradition values *method* over *outcomes*: If applying logical criteria to internally accessible evidence is the right method, we must follow it where it takes us, even if the result is total skepticism. Thus, some internalists respond that it is the cognizer's problem if her (our) beliefs fail to reach the proper standard (e.g., Fumerton, 2001). This ultimate divorce of the normative from the empirical is very odd, however. We are more confident in our beliefs about the world than we are in the theory of knowledge that motivates the internalist method. Forced to choose, we should take the more plausible of the two. We should jettison the theory rather than our everyday beliefs. We should conclude that the internalist tradition ultimately does not offer a realistic, useable concept of critical thinking.

---

[28] We will see later that neither of the two main variants of internalism – foundationalism and coherentism – can show how to justify our everyday knowledge claims by evidence that satisfies internalist constraints.

Internalists and externalists agree that for true belief to be *knowledge*, it cannot be held by lucky accident. But how should luck be excluded? The internalists try to do so by insisting on conscious reasons. The externalist, on the other hand, wants to account for the knowledge obtained by means of perception, memory, complex reasoning, and expert judgment, where no conscious reasons are available. What must be added to true belief to account for knowledge of these kinds? The externalist answer is: the *reliability* of the cognitive processes or mechanisms responsible for producing those beliefs. Luck is not a factor because the *third-person* evaluator expects beliefs formed in the relevant way to be true. In some circumstances, conscious reasoning by the cognizer may be necessary to achieve the desired level of reliability, but not always. According to Goldman (1979/1992), knowledge can be explained in terms of *reliable cognitive belief-generation processes.* Reliability of a cognitive belief-generation process "consists in the tendency of a process to produce beliefs that are true rather than false" (p.113). As noted by two of its critics (Conee & Feldman, 1998/2000), "Reliabilism is the most widely discussed contemporary epistemological theory."[29]

Internalists and externalists try to exclude luck for very different reasons. For externalists, the *purpose* of the normative evaluation of beliefs is not the fair assignment of praise or blame. Instead, the purpose is pragmatic: to determine which sources of beliefs one can *actually* trust or rely on. What matters is not inner purity from the cognizer's point of view, but whether or not she acquired beliefs in a way that would reliably lead to success from a relevant *third person* point of view. Feedback from the third person evaluator or trainer does not imply that the trainee is guilty of anything. Its intent is to improve or select rather than blame.

When a belief is produced by a reliable process, mechanism, or faculty, we may say that it is *warranted*, even if the cognizer is not aware of explicit reasons that *justify* it.[30] Even without conscious awareness of reasons, it is generally true that:

- Perception and memory are highly trustworthy under certain favorable conditions.
- Experts' judgments on matters within their domain of expertise tend to be reliable.
- Beliefs are trustworthy when they were generated by a reliable method in the past and then reliably recalled now (even if the exact reasoning is not also recalled).
- The processes used by children and animals to form certain simple kinds of beliefs are also highly reliable.

If facts such as these warrant claims to knowledge, then internalism is wrong. Two individuals might be identical in the contents of their conscious awareness, yet the knowledge or justification we attribute to their beliefs might be very different.

---

[29] Siegel (1997), among other internalists, argues that *truth* cannot be the goal of critical thinking. Since the only way we have of gaining truth is to seek *justification* by reasons for and against, the latter must be the real objective. This is internalist orthodoxy, but it is simply not the case. We can arrive at true beliefs by other processes than explicit justification., e.g., by perception, recognition, recall, relatively automatic inference, and so on.

[30] *Warrant* is Plantinga's term for whatever it is that has to be added to true belief to constitute knowledge. Sosa speaks of *aptness*. Goldman continues to use the word *justification*. Others refer to *objective justification.*

Example

MAJ Jones believes that she sees an enemy T-62 tank. Visual conditions are good, and she is positioned relatively close to the tank. MAJ Jones is highly accurate in differentiating T-62s from other kinds of tanks, and used that skill on this occasion.

LT Smith noticed the same set of features as MAJ Jones under the same conditions and arrived at the same conclusion, that the vehicle is a T-62. However, LT Smith lacks broad skill. She has some beliefs about criteria for identifying types of tanks, most of which are wrong. Her one correct rule is the one that happened to apply on this occasion. Thus, she would have identified a variety of other models incorrectly as T-62s, and would have identified most T-62s incorrectly as other kinds of tanks. She was lucky on this occasion.

We would say that MAJ Jones *knows* that there is a T-62 tank, while LT Smith does not, even though their conscious evidence and conclusions are exactly the same.

This example shows that external factors count. Knowledge involves more than the contents of one's thoughts at a single moment. Also relevant are the appropriate underlying skills, background beliefs, and history of thinking about the problem. Regardless of what cognizers think about their own ability, it is the *reality* of the underlying skills, beliefs, and history that counts, not internally accessible *beliefs*. Reflective knowledge, knowing that you know, is important in some circumstances, but it is an optional rather than a necessary part of ordinary knowledge.[31]

Externalist Definition of Critical Thinking

The goal of *justification* led to two internalist constraints: that evidence be consciously accessible evidence and also associated with a high probability of truth. What new constraints must now replace these internalist requirements? First, there must be conditions that influence the reliability of belief generation mechanisms. Under some conditions, the resulting beliefs will tend to be true; under other conditions, the resulting beliefs will tend to be false. Internalism has failed in showing that its corresponding constraint can be satisfied: that reasoning that refers only to a person's thoughts can credibly support beliefs about the external world. Externalist propose a far more reasonable empirical hypothesis, that there are facts about cognitive processes and the conditions of their use that enable us to predict the reliability of the beliefs they generate. For the support of this hypothesis, we must draw from cognitive psychology.[32]

---

[31] An internalist might respond by pointing to *occurrent beliefs* about *reliability* to distinguish MAJ Jones from LT Smith. If MAJ Jones has an occurrent belief that she is reliable in identifying T-62s, while LT Smith does not have a corresponding belief about herself, there would be an internal difference in their consciously accessible evidence after all. But it is implausible to suppose that people must have such beliefs in order to have knowledge. Moreover, beliefs about one's own reliability can be wrong. Suppose that MAJ Jones is overly humble; she believes that she is not sufficiently skilled to instantly recognize a T-62, even though she in fact is. Does MAJ Jones *know* that the tank is a T-62 under these circumstances? Internalists would say she does not, since she does not *take* the evidence to be good. Externalists would say that MAJ Jones *knows* that the tank is a T-62, but does not *know that she knows*.

[32] Of course, internalists protest that this maneuver simply begs the question by presupposing the justifiability of beliefs about the external world, e.g., those utilized in cognitive psychology. This kind of circularity is an inescapable and benign feature of knowledge. Ultimately, the third-person must also be regarded as a point of view. See the next footnote and the chapter on coherence.

Second, for an evaluation to be feasible, it must be possible to identify the conditions that affect the reliability of belief generation processes. Thus, some sort of accessibility constraint is still appropriate. Conditions that predict reliability of beliefs must be ascertainable by the evaluator. The externalist accessibility constraint is pragmatic in a way that the internalist one was not. It is accessibility to the *evaluator* that matters, and the evaluator need not be the person whose beliefs are under evaluation. The conditions of reliability need not be part of the evidence used by the critical thinker, and need not be "immediately accessible" to her conscious awareness. This is important even when the evaluator and the critical thinker are the same person. It implies that when a cognizer steps back and puts on the evaluator hat, she can profitably focus on information other than the information she used in the process of reaching her conclusion. She might consider, for example, the general reliability in the past of the thinking strategy she has just employed.

Moreover, it is the critical thinking process that is being evaluated, not the evaluator herself. Thus, she is under no obligation to provide a defense of her own views about reliability. They can be based, for example, on a plausible cognitive theory about perception, memory, or reasoning, or on systematic observation of how experts actually make decisions in a particular domain. But there is no infinite regress of demanding reasons for her judgments, reasons for those reasons, and so on.[33]

From the externalist perspective, critical thinking occurs when the cognizer adopts a third-person perspective on her own or others' belief-generating process.[34] Here then is a normative definition of critical thinking, from an externalist point of view.

---

[33] Of course, it might sometimes be legitimate to evaluate the evaluator, questioning the beliefs she used in her evaluation. But there is no threat of infinite regress and/or viciously circular justification. The answer to this is a naturalistic attitude toward epistemology (Quine, 1994) and an emphasis on overall coherence. We always take some beliefs for granted in order to evaluate others. It contributes nothing to the coherence or plausibility of these beliefs to add successive layers of redundant evaluation, in which essentially the same theories are used for the evaluation at each successive level. After the first meta-level, there is typically a point of diminishing returns. Sosa (1991) discusses the ultimate dependence of externalist models on coherence with other beliefs, as well as the reciprocal legitimation of coherence by its reliability as a process for generating true beliefs.

[34] When we speak of belief generation, what is meant here and elsewhere is generation *or sustainment*. A belief may have been acquired initially in a faulty way (e.g., by guessing), but it may then be confirmed and thus retained for more legitimate reasons.

Normative Definition of Critical Thinking #4. Externalist

Purpose	*To accept only beliefs that have a sufficiently high probability of being true.*
Constraints	*(1) Beliefs that are generated in specifiable ways under specifiable conditions have a high probability of being true.*[35]
	(2) Facts about the manner and conditions of belief generation must be ascertainable by the evaluator.
Functions	*Critical thinking is:*
	(1) the identification of facts about the manner and conditions under which a belief has been generated,
	(2) evaluation of the reliability of belief generation given the identified facts, and
	(3) acceptance or rejection of beliefs based on that evaluation.

What Does the Externalist Paradigm Look Like?

Externalism would be of little interest, even though it differs from internalism in fundamental assumptions, if it had the same implications for critical thinking theory and training. This, however, is far from the case. On the contrary, the two paradigms oppose one another on virtually every normative, cognitive, and applied issue. Let us revisit the variations in current usage left open by the consensual definition of critical thinking:

Normative issues

- *Generalizable to other intellectual products.* Internal justification, as we noted earlier, applies most directly to the ethical evaluation of actions in terms of fulfillment of internal duty. An assumption about the necessary connection between internal criteria and truth is required in order to extend internalism to beliefs. (Even less defensible assumptions about a necessary connection between inner and outer success would be required to extend internalism further, to instrumental, goal-directed actions and other intellectual products.) By contrast, externalism can be applied to any type of intellectual product (instrumental actions, inventions, stories, works of art) for which there is an identifiable criterion of external success (analogous to *truth* in the case of beliefs). Critical thinking may include monitoring and improving the conditions and methods used to construct intellectual products of any kind. The evaluator must identify facts about the manner and conditions under which the intellectual product was created, and use those facts to predict the success of the product. The following is a more general definition of critical thinking, which substitutes *intellectual product* for *belief*, and *successful* for *true*.

[35] When we speak of belief generation, what is meant here and elsewhere is generation *or sustainment*. A belief may have been acquired initially in a faulty way (e.g., by guessing), but it may then be confirmed and thus retained for more legitimate reasons.

Normative Definition of Critical Thinking #5. Generalized externalist

Purpose	*To accept only intellectual products that have a sufficiently high probability of being successful.*
Constraints	*(1) Products that are generated in specifiable ways under specifiable conditions have a high probability of being successful.*
	(2) Facts about the manner and conditions of product generation must be ascertainable by the evaluator.
Functions	*Critical thinking is:*
	(1) the identification of facts about the manner and conditions under which a product has been generated,
	(2) evaluation of the reliability of product generation given the identified facts, and
	(3) acceptance or rejection of products based on that evaluation.

- *Processes versus snapshots.* Contrary to internalism, static properties and relations among thoughts are neither necessary nor sufficient for knowledge. They are unnecessary because processes of perception, memory, recognition, and even reasoning can generate warranted beliefs in the absence of conscious reasons. Static properties and relations are insufficient for justification since a person who has all the requisite evidence in consciousness, may still not *see* its connection with the conclusion. Externalism makes the process or faculty primary, rather than the conscious events accompanying it: A belief is warranted only if it is caused in the appropriate way by a reliable process or faculty (corresponding to the first and second rows of Figure 4) under appropriate conditions (corresponding to the third row of Figure 4).
- *Domain-specific relationships.* Processes may be learned or innate, general or specific. As long as they are reliable under specifiable conditions, the input-output functions underlying a cognitive process need not be universal or a priori. For example, they may involve causal knowledge that can only be learned through experience and training in a particular domain.
- *Context-sensitive evaluation.* Externalism is, in a sense, the claim that knowledge is context-sensitive, since external factors are, by definition, the context within which conscious thinking takes place. External factors – concerning the environment, mechanisms, or processes – may make the difference between warranted and unwarranted belief. We saw that two individuals with the same *conscious* thoughts may differ in the beliefs they are warranted in accepting because they have different skills. A belief may be justified for an expert, but not for a less experienced cognizer, even if they use the same evidence. It is also possible for a belief to be warranted in a simple environment, but not warranted in a more complex situation where, for example, deception is a possibility – even though the cognizer is not conscious of the difference. Another sort of dependence on context involves stakes (DeRose, 2000). A

belief arrived at with little thought (i.e., by a process with low reliability) may be warranted in a situation where the stakes are low, but not warranted in a situation where the stakes are high.

- *Formal logic less central.* From the externalist point of view, inference is a set of input-output functions that take beliefs as inputs and produce other beliefs as outputs. The reliability of an inference process depends on the existence of an *actual* correlation between the truth of the input beliefs and the truth of the output beliefs. Thus, an inference process may be reliable even though the cognizer is not explicitly aware of any logical relations between the input and output beliefs. Indeed, an inference process may be reliable even if there *are* no logical relations (in the formal sense) to be aware of. A correlation between premises and conclusion may be due to causal relationships between facts of one type and facts of another type. It is not necessary to reconstruct everyday inferences as if they were logically valid before evaluating them. Formal logic is only one among many ways to identify reliable connections between conclusions and evidence.[36]

Cognitive processes

- *Identifying biases.* A key function of critical thinking is finding conditions under which belief generation is likely to be unreliable. Thus, by definition critical thinking involves the diagnosis of specific types of systematic mistakes (i.e., fallacies. or biases) that are likely to occur under specified circumstances. This includes one's own assumptions or interests, which may tend to distort the reasoning process.

- *Critical discussion with others.* Internalism focuses on how *individual* cognizers justify their beliefs. Interaction with others has a secondary role, only if it provides acceptable evidence for a conclusion drawn by the individual. For an externalist, by contrast, the interaction *itself* may be a reliable belief formation process. Some externalists have argued that the reliability of group or team decision making has the same status as the reliability of an individual thought process (Goldman, 1992). Dialogue can contribute *directly* to justification of the beliefs, not indirectly by adding to the evidence possessed by an individual. Exposing views to challenge reliably ensures that surviving beliefs are more likely to be true.

- *Active information seeking.* For internalism, the use of strategies in justification would introduces unfairness. It would penalize cognizers based on lack of prior knowledge and skill, factors supposedly outside of their control. But this notion of fairness is too extreme and too ill-defined. A thought in current focal awareness is supposedly equally accessible to all cognizers, but even in this case, it is not obvious

[36] An even more telling point (noted earlier) is that logical relations are not truly internal in the required sense. First, applying logical criteria requires identification of the logical form of a proposition, and that demands sensitivity to context and understanding of the meanings, substantive implications, presuppositions, and contextual connotations of the statements (Woods, 2000), as well as the "projectibility" of the predicates in the statements (Goodman, 1965). Far from being easy, re-interpreting an argument so that it fits standard logical syntax can be extremely arduous. Second, the logical principles themselves are subject to the same kind empirical pressure and revision that affects scientific theories (Quine, 1970; Everitt & Fisher, 1995). The acceptability of logical truths depends, to some degree, on how well they work in the context of our other beliefs. So, external facts are relevant in both applying and identifying logical criteria..

that mistakes are impossible (e.g., about the exact color or shape in a visual perception, or whether one belief follows logically from another). Thus, even internalists are forced to ask, *How much reflection* on items *already* in conscious awareness is called for? And different cognizers may have different degrees of skill in reflecting on their own thoughts. But if we must allow a role for effort and skill in reflection, with a concomitant risk of failure or error, we are on a slippery slope. Reflection might also help bring into focus thoughts in the periphery of awareness, or in working memory. Surely, they are accessible enough to count as evidence that the cognizer *should* take into account if relevant. But if *they* are, what about beliefs that can be retrieved very rapidly from long-term memory by reflection on items that are in working memory, e.g., by the kind of skilled memory studied by Kintsch and Ericsson? If those beliefs are regarded as accessible, what about other beliefs in long-term memory that take slightly longer to retrieve, and so on. How much effort, and how much time, is permissible for a belief to count as "accessible" to consciousness in the appropriate sense?

We can also ask the same questions for information present in the environment. Surely, it is fair to hold a cognizer accountable for evidence she can obtain visually simply by opening her eyes or looking around. What about evidence she can easily obtain by walking into the next room, asking the person at the help desk, looking in a dictionary, or conducting a critical discussion? What about information that demands a more expensive but feasible collection effort?

In sum, fairness does not require looking only at evidence that is already in "direct" conscious awareness, since such directness is an illusion. What is needed is that the cognizer have *reasonable* access to such information under the circumstances, i.e., that the cognizer be in a position to find the information without inappropriate effort and time given the *context*. Externalism rightly recognizes a wide variety of processes whose reliability can determine the warrant of the beliefs they generate.

Cognitive mechanisms

- *Reflection as an optional tool.* For the internalist, evidence of which the cognizer is aware does not justify a conclusion unless she also *sees* the link between evidence and conclusion. The requirement is needed, according to the internalist, to rule out acceptance of the conclusion for the wrong reasons. Unfortunately, the requirement cannot be met because it generates an infinite regress.[37] The intellectualist thirst for

[37] The second-order belief, about the link between first-order evidence and conclusion, is necessary for justification and so must be added to the evidence. Then there must be a third-order belief about the link between the expanded evidence set and the original conclusion, and so on. (See footnote 25.) A second kind of regress arises if the second-order belief, like other beliefs, must be supported by evidence (otherwise it might be believed by accident). This evidence requires evidence in turn, and so on. The two regresses compound one another. The second-order evidence is not justified unless the critical thinker sees how the third-order evidence supports it, and so on.

Yet another infinite regress arises if awareness of each item of evidence is regarded as part of what constitutes justification (Fumerton, 2001, p. 5), and is thus itself added to the evidence. One would have to be aware that one was aware of a particular item of evidence, aware that one was aware that one was aware, and so on. (This also interacts with the other two regresses.)

The only way for an internalist to escape a regress is to insist that awareness is *not* one of the factors constituting justification; i.e., that awareness is present simply because the factors that *do* constitute justification

self-reflection can never be slaked, and thus nothing can ever be justified. Externalism does not demand reflection on the relation between evidence and conclusion. It rules out chance acceptance of the conclusion by demanding an appropriate *causal* relation with the evidence. A warranted conclusion is one that has been generated by a reliable process or mechanism (e.g., perception, recognition, recall, or reasoning) operating upon the appropriate evidence (e.g., sensory inputs or other beliefs) as input – whether or not the cognizer is aware of the causal relationship. A belief generating process may be reliable in the absence of conscious awareness or knowledge of the causal relationships it forges between its inputs and outputs.

On the other hand, there may be contexts in which awareness of such causal relationships does add to reliability. In such contexts, the cognizer will benefit from explicitly considering the belief generating process / mechanism and its conditions of operation, i.e., by adopting a third-person, external perspective. For example, if viewing conditions are degraded or tend to produce illusions, perceptual beliefs may be subject to doubt and a second look from a different vantage point might be appropriate. If a situation is novel, then an assessment arrived at hastily might deserve more careful re-examination or discussion. In these cases, the critical thinker thinks about the performance of her own belief-generating processes and selects appropriate strategies. But critical thinking is an optional reflective attitude toward processes like perception, memory, and reasoning, which do not themselves *require* such reflection.

The introduction of the third-person, external point of view can occur at a number of increasingly general and more inclusive levels, each making decisions about the more specific level below it (Figure 6). At the most general level, long-term critical thinking habits lead to anticipation of and planning for problematic conditions before beginning an activity. Planning for an activity in turn leads processes of monitoring the activity as it occurs and choosing strategies based on monitored factors. Finally, at the most specific level choices of a strategy lead to specific actions to execute the strategy. These correspond to different levels of supervisory control. The introduction of a third-person evaluation of the belief-generating process adds a new layer to (A) specific occurrences of perception, memory, or reasoning. The new layer (B) includes (i) monitoring key factors such as perceptual conditions or the novelty of the situation, and (ii) adjusting perceptual or reasoning strategies accordingly. (B) represents the third person, external perspective on whatever specific strategy (A) is currently being executed, and thus consumes some "spare" working memory and attentional resources (Kahneman, 1976).

The new process (B) might itself invite a third-person perspective at a more general and inclusive level. B may improve reliability during some types of activities but not in others. Therefore, instead of implementing B automatically, it may pay for the cognizer to devote some thought at the start of a mission or activity to deciding what

---

(occurrent beliefs) are such that we are in fact always aware of them. But if this is so, what is the rationale for basing justification on occurrent beliefs? Occurrent (conscious) beliefs are taken to be necessary only because awareness is regarded *normatively* necessary for justification. Otherwise, it is not clear why beliefs in long-term memory would not be sufficient for justification. Yet, if awareness is necessary for justification and justification must be within our power, then we must be aware that we are aware, and we get the regress.

factors if any she will monitor and what she will do if certain conditions are observed. In planning a mission where situations are expected to be novel or where visual conditions are likely to be degraded, she may decide to monitor the degree of novelty or the quality of visual conditions. In more routine missions, it may not pay to devote cognitive capacity to such monitoring or to develop contingency plans ahead of time. In order to choose among variants of (B), therefore, there may be a new process (C), which occurs at the start of an extended activity. Decisions made at C determine the attention and working memory resources that will be consumed by B, as well as the specific strategies that will be adopted to carry B out. C also consumes some resources itself at the time it occurs, and by imposing demands on episodic and prospective memory (i.e., memory for decisions made and intentions formed early in the mission).

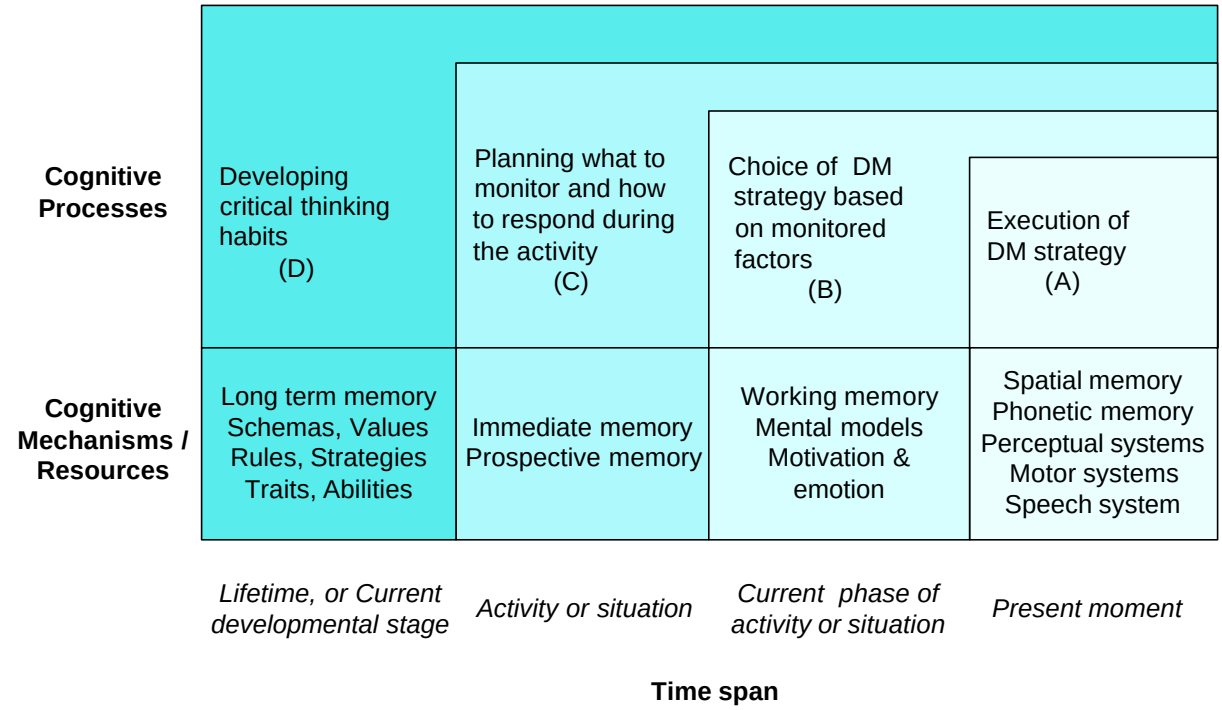


Figure 6. Points at which the external reliability of cognitive processes can be assessed, their demands on resources, and their temporal span.

The new process (C) can be utilized in each of a series of missions or over an entire career. Must C itself be evaluated for reliability at a still higher level? Is a vicious regress lurking here, where each successive layer of evaluation demands its own evaluation? The answer is no. Unlike the internalist case, the reflective processes referred to here (viz., B and C) are *optional* and at the same time, if they do occur, *feasible*. They are optional because reflection on thinking at any particular level is not a priori *required* for that level of thinking to function reliably. Indeed, a level is reached quite soon where warrant is grounded in unreflective reliability. In Figure 6, the cognizer has developed a habit (D) over the span of a career, of *reflective vigilance* – for example, reflecting at the beginning of a mission about any unusual factors that should be monitored and what should be done if they occur. Such a habit

might be one of several critical thinking *traits* which the individual inherited or developed *without* the benefit of conscious reflection or evaluation.

The reflective processes (B and C) are feasible because they do not consume unlimited cognitive resources (see second row of Figure 6). B demands some "spare" working memory and attention since it occurs in parallel with lower level cognitive processes (A). C demands only occasional working memory and attention during the planning phases of activities, and some immediate or prospective memory to keep track of the plan during the activity itself. D resides in long-term memory and other persisting structures and thus requires no dynamic cognitive resources. Therefore, while internalists seem committed to endlessly escalating, redundant levels of introspection, externalist posit realistic cognitive processes (B and C) and traits (D). These draw on different types of resources and contribute in diverse ways to the reliability of outcomes over different temporal spans (Figure 6).

- *Effort graded to conditions.* Internalism insists that effort be expended to scrutinize *every* habitual or automatic belief. Externalism reminds us that in many cases such effort may be counterproductive, since (i) capacity is not available to accomplish it, (ii) even if capacity were available, automatically generated beliefs and actions are often trustworthy, and (iii) even when they are not trustworthy, there may be little or nothing at stake. This is precisely the kind of decision (C) about what to monitor and what to do, that might be made at the onset of an activity (Figure 6). As we saw in the previous example, critical thinking should *vary* in effortfulness depending on both the costs of errors and the novelty of the situation (which determines the likelihood that underlying assumptions may turn out to be wrong). Externalist critical thinking is adaptive precisely because of the possibility of external, third-person evaluation.

- *Traits and attitudes.* According to the original, and narrowest, externalist view, the warrant for a belief derives from the reliability of the *process*, or input-output function, used to generate the belief (Goldman, 1992, p. 115). Externalists have more recently begun to focus on the reliability of more persisting faculties or traits (sometimes called "virtues") as the basis for warrant (e.g., Goldman, 1989; Sosa, 1991, chapter 8; Plantinga, 1993b). The externalist's motivation for introducing reliability in the first place was to exclude the role of luck: Beliefs that are true because of chance are not justified. But temporal scope is a key parameter in identifying the role of luck. If a judgment is produced by a reliable cognitive process, then its accuracy is not due to luck. On the other hand, in a broader perspective, the acquisition of the cognitive process itself might have been due to chance; had another process been used, as it easily might have, the judgment would have been false. So, from this broader perspective, the truth of the judgment would be a matter of luck after all. But suppose the cognitive process was explicitly chosen because of its reliability. Or, alternatively, suppose that the cognitive process resulted from a skill that was developed (or inherited) because of its tendency to produce reliable processes. This extends the temporal scope within which the role of luck is excluded as a cause of the accuracy of the judgment. The broader and more persistent the *skill* underlying a judgment, the less the truth of that judgment can be attributed to luck. The cognizer with a persisting skill is entitled to more inclusive trust from a third-

party evaluator. The result will be a reduced requirement for monitoring the cognizer's performance.

Example

MAJ Jones correctly identifies a vehicle as a T-62 tank. She is reliable when differentiating different types of tanks, so her conclusion is not correct due to luck and seems to be a justified assessment. However, MAJ Jones is not reliable when differentiating different kinds of trucks. Thus, in the larger picture, the accuracy of her belief that the vehicle is a T-62 is due to luck after all: Had there been a truck instead of a tank, she would have identified it incorrectly.

MAJ Jones' superior COL Black knows this, and has to consider the type of vehicle MAJ Jones claims to see before deciding whether or not to accept her assessment. The warrant for MAJ Jones' belief that the vehicle is a T-62 would be enhanced if she possessed a broader set of related skills. And the effort required from her superior COL Black would be reduced.

If a cognizer's accuracy is highly variable and dependent on specific conditions, extensive monitoring and correction will be necessary. A third-party evaluator may be unable to determine whether to trust the cognizer's conclusions without expending more effort than it is worth. An enduring faculty or trait that leads to consistently appropriate processes in a wide range of situations is far better.

Externalists have tended to choose a level of generality which they regard as *the* correct focus for evaluation of reliability. Both cognitive processes and enduring traits have been proposed. We suggest, as Figure 6 illustrates, that there is no one right level. Warrant depends on reliability at a variety of different levels. A judgment that the enemy will attack in the north is the outcome of a process of reasoning (A), processes for selecting and executing reasoning strategies (B), processes for allocating cognitive resources (C), and a persisting trait to think critically (D). The wider the temporal scope over which reliability is assured the better, and the more generally accurate one's judgments will be. Critical thinking is a reflective process that can occur at a variety of supervisory levels. The more reliable the underlying processes, the less frequently critical thinking needs to occur and the higher the level at which supervision takes place.

How Do Views From Inside and Outside Differ?

In the chapter before last, we reviewed the definitions of critical thinking that have appeared in the literature and analyzed the significant variations among them. Divergences were classified into normative, cognitive process, and cognitive mechanistic categories, and several dimensions were identified within each category. We have now shown that two high-level paradigms – internalist and externalist –account for *all* the variation in those definitions. Internalist and externalist paradigms differ on *every* dimension along which we found variation to exist at all, as shown by the following tables.

Normative	**Internalist**	**Externalist**
Objects to be evaluated	Only beliefs and ethical actions are evaluated	Any intellectual product can be evaluated, including recognition, instrumental actions, stories, art, etc.
	Evaluate static snapshot of consciousness	Evaluate reliability of cognitive processes
Criteria of evaluation	Apply universal standards	May be domain-specific
	Context-independent	Context-dependent
	Logic is privileged	Logic is not privileged

Cognitive Process	**Internalist**	**Externalist**
Strategies	Follow rules for correct reasoning	Identify biases and fallacies in own reasoning
		Expose own views to challenge and to opposing positions
	Only information currently in mind is considered	Actively seek information from long term memory or by data collection

Cognitive Mechanism	**Internalist**	**Externalist**
Conscious "will"	Reflection is necessary for justification	Reflection (i.e., critical thinking) is optional
	Effort required to uproot all habitual or automatic beliefs	Effort varies with familiarity and stakes in the context
Character	Evaluation focuses on individual occasions and is always required for justification	Enduring traits and aptitudes enhance reliability, hence, increase warrant and reduce the need for evaluation on a particular occasion

Although specific approaches to critical thinking tend to mix and match from the two columns, internalism and externalism seem to bound the possible positions on the key issues. The tension between first-person (internalist) and third-person (externalist) points of view helps us understand otherwise puzzling differences among critical thinking theories.

Differences for Training

The internalist motivation was to provide norms that can be *followed* by cognizers, because they refer only to evidence of which she is directly aware. Thus, failure to follow them is legitimately subject to blame. Externalist norms, on the other hand, may "unfairly" punish the cognizer for failing to take account of information that was not available to her. More importantly, how can externalist principles guide action if they refer to conditions outside awareness? Will the externalist paradigm make any sense as a basis for training critical thinking?

In fact, externalist norms make *more* sense for training than internalist ones. On the one hand, as we have already seen, internalist norms are unrealistically difficult. They require conscious reasons in cases (perception, memory, expert judgment, inference based on large stores of information) where conscious reasons are not necessary. Many of our most plausible beliefs cannot be justified in an internalist framework. Training to be reflective about such beliefs is likely to disrupt rather than improve performance.

On the other hand, internalist norms are unrealistically easy on the cognizer in other respects. As we saw, they make no provision for *strategies* to uncover new information by shifting attention in working memory, retrieval from long-term memory, discussion with others, or exploration of the environment. The dynamic aspect of critical thinking is completely suppressed and therefore cannot be addressed by training. From the externalist perspective, on the other hand, the purpose of training is to improve the reliability of the processes and strategies by means of which cognizers generate beliefs (or other intellectual products). This includes training cognizers to acquire strategies for effective marshalling of relevant information wherever that information may be found. More generally, the objective of training is not simply to teach critical thinking, but to inculcate habits and attitudes, i.e., to produce *critical thinkers*.

6. TACTICS: A MENU OF MID-LEVEL PARADIGMS

In the previous chapter, we showed that two high-level paradigms – internalist and externalist –account for virtually all the variation in current definitions of critical thinking. But there is probably no extant position that is purely internalist or purely externalist. Instead, there is a set of distinct *middle level* paradigms such as formal logic, informal logic, decision theory, dialogue theory, rhetoric, naturalistic decision making, bounded rationality, communication studies, and others, which draw in various ways from the high-level paradigms, internalism and externalism. Textbooks, theories, and training techniques for critical thinking draw in turn from these middle-level paradigms. It is often unclear how ideas borrowed from the different mid-level paradigms mesh with one another, and sometimes they actually seem to be in conflict. As a result, critical thinking theory and practice gives the appearance of unsystematic eclecticism. Applying the results of the last chapter to the more familiar middle-level critical thinking paradigms can bring some order to this chaos. In this chapter, therefore, we take a look at how the important mid-level paradigms vary on the spectrum of views bounded by internalism and externalism.

What Features Distinguish Mid-Level Paradigms?

Different theories of critical thinking agree that a necessary function of critical thinking is to construct reflective arguments about the normative adequacy of intellectual products. The conclusion of the higher level argument serves the *purpose* of critical thinking, to determine whether or not the intellectual product is normatively adequate. The evidence for that conclusion is how well relevant facts about the intellectual product fit evaluative criteria. Internalist and externalist paradigms agree at this abstract level. However, as Figure 7 shows, they differ rather profoundly on how the components of the higher-level argument are to be fleshed out.

(In addition to evidence and conclusion, Figure 7 depicts two other argument components, based loosely on Toulmin (1958): *warrant* and *backing*. The warrant of an argument explains why we are entitled to infer the conclusion from the evidence. According to Toulmin (p. 98), it corresponds to "practical standards or canons of argument" that license the inference step from evidence to conclusion. The *backing* of an argument, as noted earlier, explains where the warrant came from and, according to Toulmin, "why in general this warrant should be accepted as having authority" (p. 103). We take backing to be the biological, social, and/or cognitive *source* of the warrant. The backing in a critical thinking argument about an intellectual product includes the processes of inheritance, maturation, individual learning, thinking, theorizing, argumentation, and/or cultural practice that resulted in the adoption of specific criteria of normative adequacy. Thus, the backing explains why the warrant is accepted as having authority.)

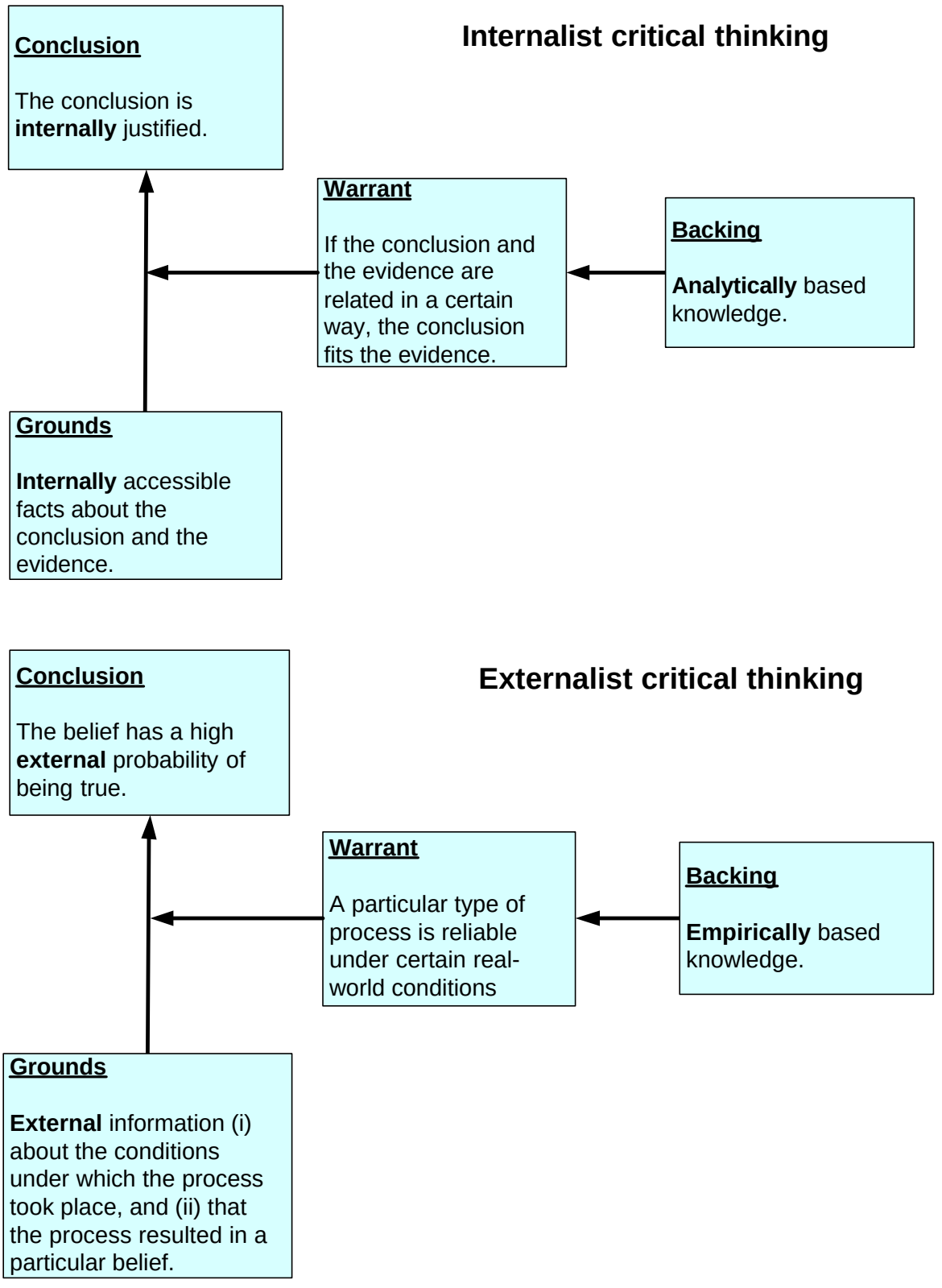


Figure 7. Critical thinking as a reflective argument about an intellectual product, from the internalist perspective (top) and externalist perspective (bottom)

Two of the components in Figure 7 are particularly useful for locating the fault lines that divide mid-level critical thinking paradigms. Thee key differences pertain to the *grounds* and the *backing*, respectively, of the higher-level critical thinking argument:

1. *Grounds*: Are the facts that serve as criteria *internal*, i.e., must they be facts to which the cognizer has conscious access and can readily form beliefs about, or can they be *external*, i.e., may they include facts about the real world or cognitive mechanisms of which the cognizer may be unaware?
2. *Backing*: Is the association between facts and normative adequacy determined in relatively *analytical* manner (e.g., by supposedly *a priori* methods such as logical intuition, inference from abstract principles, or mathematical calculation from general assumptions) or *empirically* (e.g., by observation of successful strategies and/or scientific models of cognition)? Backing is analytical if it accounts for the origins and authority of warrants by means of so-called first principles. Backing is empirical if it accounts for the origins and authority of warrants in terms of contingent facts.

We can classify theories of critical thinking in terms of: (1) application of criteria to internally accessible information only versus application to any facts whether internal and external, and (2) predominantly empirical versus predominantly analytical backing for the criteria. Crossing these two dimensions yields a taxonomy with four different approaches to critical thinking:

- *Engineering models*: External grounds, analytic backing
- *Formal models:* Internal grounds, analytic backing
- *Informal models*: Internal grounds, empirical backing
- *Naturalistic models*: External grounds, empirical backing

Examples of mid-level paradigms that fall roughly into these four categories are shown in Table 1. We will summarize briefly some characteristics of these four divisions in this chapter. Any pigeon-holing of methods is likely to be oversimplified, and this one is no different.

Table 1. A taxonomy of major positions on critical thinking, classified on two dimensions. *Backing* refers to the source of norms. *Grounds* refers to the types of facts to which the norms are applied.

	Internal Grounds *Current contents of consciousness*	**External (as well as internal) Grounds** *Any relevant facts*
Analytical Backing *Reasoning and intuition*	**Formal models** • Deductive logic • Probability / Decision theory	**Engineering models** • Constrained optimization • Cost-benefit analysis
Empirical (as well as analytical) Backing *What proficient people actually do; record of success*	**Informal models** • Informal logic • Dialogue theory$_1$ • Rhetoric$_1$	**Naturalistic models** • Bounded rationality, Adaptive decision making • Expert performance, Naturalistic decision making • Dialogue theory$_2$ • Rhetoric$_2$, Communication

Analytical Backing

Engineering. Engineering approaches develop problem-solving and decision making methods without reference to the cognitive processes that people use (analytical backing) or the beliefs they hold (external grounds). Mathematical models are developed and then used to identify optimal responses to objectively measured environmental conditions, constraints, and goals. Despite the lack of concern with human cognition, such models of ideal performance have been used as benchmarks for evaluating *human* cognitive performance by incorporating the predicted effects of human capacity constraints on ideal performance (e.g., Rubinstein, 1998).

Formal. Often the required objective measures are not available to serve as grounds. In this case, another analytical approach, represented by formal logic and decision theory, comes to the rescue. These paradigms also use analytical methods to develop idealized models, but the models are applied to subjective judgments rather than objective measures. They are concerned with whether the beliefs and subjective preferences of an individual cognizer fit appropriate criteria of formal consistency and internal coherence. Formal logic, for example, is concerned with whether an inference is valid, not whether the premises are true. Decision theory is concerned with whether probabilities assigned to various propositions satisfy axioms of probability theory, not whether they reflect actual relative frequencies of events in the world. In true internalist fashion, what the cognizer doesn't know won't hurt her as long as her beliefs and preferences conform to appropriate formal constraints. Unlike engineering models, formal

models do not take into consideration human capacity limitations (unless the decision makers happen to have explicit beliefs about them). Nevertheless, they are frequently used as benchmarks for evaluating how people actually reason and make decisions. The verdict is typically quite negative (see discussion by Cohen, 1993b).

Empirical Backing

Informal. Critics argue that it is not enough to supply subjective inputs to analytical algorithms. The algorithms themselves must more closely approximate the way humans think. Formal models fail to capture the flexibility and richness of real-world reasoning, especially by experienced practitioners in complex domains (Klein, Orasanu, Calderwood, & Zsambok, 1993). Empirical approaches attempt to remedy this shortcoming by developing criteria with at least some empirical backing. They adopt the principle that "questions about how we actually arrive at our beliefs are …relevant to questions about how we ought to arrive at our beliefs" (Kornblith,[38] 1994, p. 3). They draw on *real* cognitive processes to determine how thinking should be done. Empirical approaches differ, however, in their attitude toward grounds.

Informal logic aims to study real-life argumentation as opposed to formally rigorous logic. It attempts to characterize the way people ought to evaluate real-life arguments based in part on the way they actually do so. It looks at how people reason in settings ranging from everyday life to political debate to specialized professions and technical and scientific discourse. The principles of normative adequacy come from a mix of analytical considerations, on the one hand, and observation of and intuitions about practice, on the other. Even though the source of its norms is (in part) empirical, however, the facts to which they are applied are highly constrained in practice. The concrete context of an argument is largely ignored in examples given in informal logic textbooks (as noted by Walton, 1998). Norms apply to statements that are made and relations among those statements in a way that is reminiscent of traditional formal logic. The features attended to are general and abstract (Are the premises *acceptable*? Are they *sufficient* to establish the conclusion?). The norms apply to assertions rather than beliefs, but no features of those assertions or their external context (i.e., pertaining to their nature as speech acts) are used in the evaluation. For all intents and purposes, therefore, the norms can be applied directly to beliefs. The grounds treated in informal logic are essentially internalist.

Dialogue theory can be considered part of informal logic, but for our purposes it is more convenient to deal with it as a separate development. (In fact, it is also influenced by work in rhetoric and communication theory.) First, dialogue theory is more empirical and less analytical than (the rest of) informal logic in its exploration of real-life argumentation, identifying, for example, different types of dialogues in which argumentation takes place and which serve as the sources of different types of norms. Dialogue theory shifts the focus from static relations among beliefs to interactive processes of argumentation. Second, dialogue theory admits external facts into the evaluation of argumentation, in particular, facts about the concrete context and the purposes of the participants that should lead them to adopt one dialogue type rather than another.

Since both rhetoric and dialogue theory allow both internalist and externalist grounds, why not classify them under the heading of *external (as well as internal) grounds* in Table 1? To do so would obscure the fact that the two sorts of grounds are not indiscriminately mixed, but function in highly distinguishable ways in both rhetoric and dialogue theory. It is more

[38] Kornblith refers to this view as a form of *naturalized epistemology*.

illuminating, therefore, to discuss rhetoric and dialogue theory from the point of view of both internalist and externalist treatments of grounds.

Dialogue theory refers to two people engaged in an overt verbal exchange over a period of time. This public character, as well as the temporal extension, makes dialogue theory externalist. However, as in the case of informal logic, we have to ask what aspects of the external process are essential to the evaluation over and above what is represented in the belief systems of the participants. Dialogue theory is internalist to the extent that it focuses on internal conformity of the verbal exchange to the norms of a particular type of dialogue. If the participants in a dialogue are aware of the norms and how to apply them, and the facts to which they apply are always accessible to awareness, to that extent the participants cannot be mistaken about whether or not they are justified, and the relevant norms are internalist. Several features of dialogue theory in fact share the spirit of the internalist paradigm: First, the norms are (usually) applied only to facts that are *known* to one or both of the participants. These facts might include, for example, the issue being discussed, the beliefs to which the participants have committed themselves, and the previous move or statement in the exchange (Walton, 1999). Second, the evaluation focuses on achieving proximal (i.e., internal) objectives that are associated with a particular type of dialogue rather than on distal (i.e., external) objectives. For example, the goal of a persuasion dialogue is to resolve a conflict of opinions with another person; the goal of an inquiry is to accumulate a store of mutually acceptable propositions; the goal of a negotiation is to resolve a conflict of interests (Walton, 1999). Achievement of these objectives can be defined in terms of beliefs or intentions that each of the participants accepts, without reference to changes in external conditions. Third, according to Walton, each dialogue type is a "conventionalized" normative framework for evaluating the appropriateness of the argumentation that occurs within it. That is, argumentation techniques are evaluated in terms of their fit to (internal) expectations about a particular type of dialogue, not in terms of the external likelihood that an argumentation technique will in fact achieve the goal of that specific dialogue. In sum, a major component of the evaluation process described in dialogue theory takes place *within* the sphere of beliefs (or acceptances of propositions) of the participants in the dialogue, just as for traditional internalist models evaluation takes place within an individual mind. At this level of analysis of argumentation, dialogue theory abstracts away from actual, concrete objectives of the participants and facts about the situation (e.g., the need to accomplish a mission with a certain period of time). From this internalist point of view, the purpose of the argument is to arrive at beliefs that are *justified* within the context of that dialogue type, not with respect to success in an external task.[39] Because of their focus on internal conventionalized norms and proximal objectives, dialogues can be viewed as self-enclosed games. As we shall see, however, both dialogue theory and rhetoric have an important and quite distinct externalist components as well.

Naturalistic. Naturalistic models complement informal models by looking more broadly at the context in which thinking and decision making take place, whether that context is understood by the cognizer or not. Proficiency is not measured by internal conformity to rules, but by success or failure in achieving actual goals under actual constraints. The aim is to identify strategies that reliably achieve objectives under varying external conditions, which include

[39] One of the early contributors to dialogue theory, Hamblin, (1970, p. 255) says, "In our present discussion we shall not be concerned to consider any contact of the dialogue with the empirical world outside the discussion-situation."

properties of the task, the environment, and the cognizer (e.g., degree of experience). Because they look beyond the internal representations of the cognizer, naturalistic models can ask, as informal models cannot, *how* cognizers represent the external environment, and can identify discrepancies between internal representations and reality. They can also study the difference between experts and novices in creating successful representations of underlying structure in a particular domain of knowledge.

Despite these common interests, naturalistic models adopt different tactics. Bounded rationality (Simon, 1997) and adaptive decision making (Payne, Bettman, & Johnson, 1993) typically develop mathematical models to fit the data obtained in experimental studies of human decision processes. These models are used to predict how human cognitive strategies perform as external variables change, such as workload, number of options, and number of possible outcomes or goals. By contrast, workers in expert performance (Ericsson, 1996) and naturalistic decision making (e.g., Klein et al., 1993) often (though not always) emphasize ecological validity and therefore adopt a less experimental, more observational approach to empirical backing, and a less mathematical, more qualitative approach to modeling thinking strategies. Either of these approaches can be applied to team and organizational, as well as individual, decision making. From the externalist point of view, team decisions can be evaluated *directly* in terms of the reliability of team processes leading to the decision. Such an evaluation may be, but does not have to be, based on separate evaluations of decisions by individuals in the team.

We have classified dialogue theory and rhetoric as both externalist and internalist. The dual nature of rhetoric is based on the emphasis within that field on two quite different aims of normative evaluation. The first aim (rhetoric_1) is to evaluate argumentation techniques in terms of their suitability for *rational persuasion* of an ideal audience. The second aim (rhetoric_2) is to evaluate argumentation techniques in terms of *actually persuading* a specific real audience. Striking a balance between these two aims is a difficult issue for those who study rhetoric. The result, however, is that rhetoric has a foot in both camps: The first goal (rational persuasion of a hypothetical ideal audience) is internal, and resembles the approach taken in informal logic. From the point of view of rational persuasion, what matters is the internal fit of the argument to appropriate norms. The second goal (actual persuasion of a real audience) is external. From the point of view of actual persuasion, unless the argument has been properly adapted to the beliefs, proclivities, and culture of the real-life audience that has been targeted, it will be objectively unlikely to succeed with that audience. Rhetoric in this sense is part of the naturalistic study of expertise in the use of words, imagery, or other tools for persuasion.

The problem with the dual internal and external character of rhetoric is that the two points of view are not coordinated. It is easy to imagine the two goals – to convince an ideally rational audience and to persuade an actual real-life audience – to be in conflict, for example, if appeal to emotions would sway the real audience, but would appear irrational to the ideal audience. No clear guidance is provided for reconciling them. Dialogue theory, on the other hand, provides a more coherent bridge between the internal and the external. Dialogue theory is more general than rhetoric, applying to forms of interaction with purposes other than persuasion. Dialogue types, for example, include persuasion, negotiation, deliberation, inquiry, and information seeking, as well as their subtypes. Each type of dialogue has different rules or norms, which are internal to the extent that they apply to beliefs and interests of which the participants are aware. But there are a number of important externalist features in dialogue theory, which complement rather than conflict with the internalist aspects. First, a particular type

of dialogue is adopted in order to achieve external goals of a task or activity, and this choice therefore reflects an external point of view. The test of appropriateness is external probability of success in the relevant circumstances. Second, dialogue theorists emphasize that the "conventionalized" norms inherent in each dialogue type reflect shared expectations. But this is not incompatible with the supposition that they also reflect objective correlations between various argumentation techniques and the achievement of the goal of that particular dialogue type. Shared expectations should ultimately by shaped by such external facts. Third, the external context has a strong influence on how the internal norms are applied. For example, the weight of evidence required to establish a conclusion will be less in a deliberation dialogue, where an action must be chosen in a timely manner in the face of incomplete information, than in an inquiry, where conclusions must be arrived at carefully over time so that they will not have to be retracted later. Fourth, in some types of dialogue the rules and norms extend to *tacit* beliefs and interests, which may not emerge into awareness until the dialogue itself elicits them (Walton & Krabbe, 1995). The appropriateness of adopting different kinds of dialogue may depend on the value of knowledge that is not yet explicitly "known" to either party! The union of internal and external factors in dialogue theory is complex and subtle. It will provide the basis for an integrated theory of critical thinking to be described later.

What Is the Role of Arguments?

Is Critical Thinking the Same as Reasoning?

Arguments are a central concept in critical thinking, at two levels: First, as we just saw, the evaluation of an intellectual product (e.g., a belief) can be framed as a higher-level argument about the normative adequacy of that product (Figure 7). Second, the intellectual product itself may have been the conclusion of an argument. In that case, critical thinking becomes a higher-level argument about the normative adequacy of a lower-level argument. For internalists, critical thinking focuses on arguments because they are the premier form of belief generation (unless they are self-evident, all beliefs must be justified by explicit reasons). This brings us to a key point: For internalists, the distinction between lower and higher level arguments tends to blur and collapse. Unless the cognizer is aware of a particular consideration, it is not relevant to justification of *her* beliefs. Thus, if anything turns up in the *higher-level* argument that bears on whether or not the cognizer is justified in accepting a belief, she must be aware of it. Thus, it must also be included in the *evidence* of her original, *lower-level* argument, or else that argument is incomplete. There is no distinct outside perspective on justification for the higher-level argument to adopt; thus, the usefulness of a higher-level argument to the cognizer is in doubt. That is why for internalist-influenced approaches, critical thinking is in effect *equated* to reasoning in general. Since a justified reasoner must already be reflective, i.e., aware of the evidence, the conclusion, and the link between them, a higher-level argument is already implicit in any legitimate first-level argument. Critical thinking is omnipresent, and therefore adds nothing special.[40]

---

[40] The points made in this paragraph are related to an objection to Toulmin's theory that centers on the difficulty of making a principled distinction between warrants and linked premises (e.g., see Johnson, 1996, chapter 7; Freeman, 1991; also, Toulmin, p. 99). The problem is, Why shouldn't warrants be included as part of the evidence in the argument, rather than treated as a distinct component? As Freeman (1991) points out, when premises are *linked*, each premise is required for the effectiveness of the others. Thus, each premise "licenses the inference" from the other premises to the conclusion. But this exactly fits Toulmin's definition of *warrant*.

For externalists, on the other hand, arguments are only one type of belief-formation process. When a belief is acquired, the higher-level argument evaluates the reliability of the process by which the belief was formed. If that process is perception or recognition , this external perspective is clearly distinct from the cognizer's first-person perspective, within which the belief may have emerged automatically. Since reflection is not a necessary accompaniment of the lower level process, it is in a position to add something useful. Even when reasoning is the process responsible for a belief, externalists insist that reasoning does not *require* awareness of the link between a conclusion and the reasons for it. The reasoning process that takes the evidence as input and produces the conclusion as output may be a black box to the cognizer herself (e.g., it may operate by constraint satisfaction in a neural network). Reflection on the reliability of that process can only take place from the third-person perspective.

It is important to realize, however that even if the first-level argument does involve reflective reasoning, an external viewpoint is useful. It adopts a *more general* perspective than the cognizer does in her first and second person roles. The higher-level argument inquires into the reliability of a process *type* of which the current reasoning is only one instance. It thus provides insight not available from the first-person point of view.

Arguments in Internalist Paradigms

Table 2 summarize how arguments are viewed in several mid-level internalist paradigms. As the first column shows, in each case the purpose of a critical thinking argument is to establish that a conclusion is justified by a lower level argument. For formal logic, justification involves rules that guarantee transmission of truth from premises to conclusion. For informal logic, dialogue theory and rhetoric, justification also involves voluntary conformity of inferences to rules. Since such conformity is necessary for justification, the cognizer must be aware of it. Thus, in each internalist paradigm, the rules must be *internalized* norms.[41]

One suggestion (Toulmin, p. 100) is that warrants can be distinguished from premises by virtue of being *implicit*. The problem is, from the internalist point of view, if a warrant is relevant at all to the argument, it must be consciously available to the cognizer, i.e., readily capable of becoming explicit. And as Freeman says, making a proposition explicit should not magically change its *function* from warrant to premise. Internalism appears to leave no room for a proposition that is "outside" the argument (i.e., neither premise nor conclusion) but nonetheless relevant to the argument. Thus, it cannot recognize warrants as a distinct category.

On the other hand, an externalist perspective can easily make sense of warrants as a special argument component distinct from premises. We define warrants as the *cognitive processes*, i.e., *input-output functions*, that are responsible for the generation of the conclusion from the evidence (the top row in Figure 4). The *backing* component in Toulmin's scheme can be distinguished from evidence for the warrant by a similar move. Backing can be defined, externally, as a more temporally extended process, which serves as the *source* of the cognitive process that serves as the warrant. Backing might involve cognitive processes like learning, analytical reasoning, or empirical research, as well as non-cognitive processes like evolution and maturation,. The key point is that warrants and backing are both *processes*, hence, they are different kinds of things from premises. They are not beliefs, although a cognizer may have reflective beliefs about them.

[41] As we saw in footnote 40, the cognizer must be aware of the norms that higher level arguments apply, if they are relevant to justification. But adding the norms to the evidence changes the original lower-level argument. We have an infinite regress here, since new norms are now needed to evaluate the revised lower-level argument.

Table 2. *Internalist* views of conclusion, evidence, and criteria of higher-level critical thinking arguments.

Field	**What is conclusion = purpose of evaluation?**	**What is evidence = facts about lower level argument that are considered?**	**What is warrant = conditions under which the lower-level argument is acceptable?**
Formal Logic	Show that the conclusion of the lower level argument must be true given its premises.	Argument = "Any group of propositions … one of which [the conclusion] is claimed to follow logically from the others [the premises]" (Soccio & Barry, 1992; in Walton, 1996b)	(i) Premises are logically consistent. (ii) Inference is formally valid: each step follows by syntactically defined rules.
Informal logic	Show that the conclusion of the lower level argument is justified by its premises.	Argument = "A set of claims that a person puts forward in an attempt to show that some further claim is rationally acceptable. Typically, people present arguments to try to persuade others to accept claims" (Govier, 1997, p. 2)	Reasons are (i) rationally acceptable to participants, and are (ii) relevant and (iii) sufficient to establish the conclusion. (iv) Argument is not fallacious.
Dialogue theory$_1$	Show that both participants have collectively achieved the goal of the dialogue, and that one participant has won the dialogue.	Argument = "…reasoning used in various types of dialogue … (Walton, 1996b: p. 11) Dialogue = "…a goal-directed conventional framework in which two speech partners reason together in an orderly way…. Each type of a dialogue has distinctive goals as well as methods that are used by the participants to achieve these goals together." (Walton, 1998: p. 3)	(i) Sequence of exchanges between participants conforms to rules or norms associated with each stage of that type of dialogue. (ii) Participants avoid inappropriate shifts between different types of dialogue. (iii) One participant wins according to the rules by defeating challenges to her viewpoint.
Rhetoric$_1$	Show that audience would be rationally persuaded to accept the conclusion.	Argumentation = use of "discursive techniques allowing us to induce or to increase the mind's adherence to the theses presented for its assent" (Perelman & Olbrechts-Tyteca, 1969: p. 4)	Speaker uses techniques that would rationally persuade a hypothetical "universal" audience, composed of all reasonable and competent people.

The definition of *argument* varies in an interesting way among the different internalist paradigms. In ordinary speech, the word *argument* can take different meanings. For example, an argument may be an impersonal *proof*, a rational *discussion*, or a highly personal *quarrel*. The

range of information that we would consider relevant to the "argument" increases as we go from proof to discussion to quarrel. In a similar way, even though (for internalists) information pertaining to justification must be consciously known by the cognizer, the scope of that information, and the definition of argument, becomes more inclusive as we move from formal logic to informal logic to dialogue theory to rhetoric. As shown in Table 2, at one extreme *formal logic* evaluates first-level arguments in terms of the syntactic relations among a set of statements (consistency of the premises and entailment of the conclusion by the premises). No other information is relevant. *Informal logic* takes into account the subjective motive of the participants in an argument (to persuade), the subjective acceptability to the participants of the premises, and the subjective degree of support that participants believe the evidence lends to the conclusion (where this support need not be based on a syntactically defined relationship).[42] Neither dialogue theory nor rhetoric focuses on arguments as the basic unit of study. Each places arguments in a larger communicative context. *Dialogue* theory moves beyond a static representation of the evidence and conclusion, taking into account the *history* of moves in a temporally extended process. It also takes into account various possible subjective goals of the participants (not just persuasion) and the corresponding types of dialogue. Finally, *rhetorical* models take into account information about the beliefs, interests, and cultural context of the audience.

Outcome and Process in Externalist Paradigms

The purpose of critical thinking in the externalist mode is not internal justification but achievement of external objectives (see first column of Table 3). Engineering models are analogous to formal logic: Both seek to guarantee a uniquely correct outcome (valid conclusion or optimal adaptation, respectively) from a given starting point (premises in informal logic and assumptions and constraints in engineering models). Both are based on analytically derived principles. The difference is that engineering models use external facts as inputs to the solution.

Bounded rationality and naturalistic decision making do not use analytically derived principles. They are based on empirical observation of actual decision makers. For this reason, they are analogous to informal logic in their attitude to the purpose of critical thinking: Each sets a less ambitious goal in comparison to formal logic and engineering models. For informal logic, the purpose is justified belief, i.e., *probable* rather than guaranteed truth. For bounded rationality and naturalistic decision making (and, plausibly, for rhetoric), the purpose is to find a solution that is "good enough" even if not optimal. The difference is that bounded rationality and naturalistic decision making utilize external facts as inputs, while informal logic does not.

[42] Of course, as discussed earlier, these internal, subjective assessments are assumed to *correlate* with objective probability of truth.

Table 3. *Externalist* views of conclusion, evidence, and criteria of critical thinking arguments.

Field	**What is conclusion = purpose of evaluation?**	**What is evidence = facts about lower level argument that are considered?**	**What is warrant = conditions under which the lower-level argument is acceptable?**
Engineering	Select the combination of parameter values that maximizes an objective function.	Any facts at all about available options, objectives, and constraints, as long as the information is precise enough to be susceptible to modeling.	Selected option is best given the assumptions that have been made about the problem.
Bounded rationality	Select an option that is "good enough," i.e., meets or exceeds specified criteria.	"…(a) seek to determine empirically the nature and origins of values and their changes with time and experience; (b) seek to determine the processes, individual and social, whereby selected aspects of reality are noticed and postulated as the "givens" (factual bases) for reasoning about action; (c) seek to determine the computational strategies that are used in reasoning…; and (d) seek to describe and explain the ways in which nonrational processes (e.g., motivations, emotions, and sensory stimuli) influence the focus of attention and the definition of the situation that set the factual givens for the rational processes." (Simon, 1997: p. 368)	Decisions are made "in a way that is procedurally reasonable in the light of the available knowledge and means of computation" (Simon, 1997: p. 369).
Naturalistic decision making	Achieve individual or organiza-tional objectives.	"…asks how experienced people, working as individuals or groups in dynamic, uncertain, and often fast-paced environments, identify and assess their situation, make decisions and take actions whose consequences are meaningful to them and to the larger organization in which they operate." (Zsambok, 1997, p. 5)	Strategies are characteristic of proficient decision makers in the relevant field.
Rhetoric_2	Persuade a real audience to accept belief.	Argumentation = use of "discursive techniques allowing us to induce or to increase the mind's adherence to the theses presented for its assent" (Perelman & Olbrechts-Tyteca, 1969: p. 4).	Speaker uses techniques likely to persuade a specific audience to adopt the point of view being presented.

The target of critical thinking evaluation in the externalist mode need not be an argument. In fact, engineering approaches usually by-pass human cognitive and communicative processes altogether. They directly assess alternative solutions or products rather than providing an evaluation of arguments for and against the solutions. Human cognitive processes can be evaluated within the engineering paradigm by comparing their expected outcomes with the results of the optimal algorithms modulo human resource limitations. Empirical approaches, on the other hand, look directly at the reliability of a *human process* for achieving "good enough" outcomes. Thus, bounded rationality, naturalistic decision making, and rhetoric are analogous to dialogue theory in their emphasis on process rather than outcome. For bounded rationality and naturalistic decision making, the processes to be evaluated are *cognitive*. For rhetoric, the relevant processes are *social* and *communicative*. Argument is prominent among the tools of persuasion that rhetoric focuses on. Yet even in the case of rhetoric, the relevant process may involve non-argumentative elements such as stories or visual imagery.

There is potential for complementarity between dialogue theory and bounded rationality / naturalistic decision making. On the one hand, dialogue theory provides an internal characterization of cognitive / social processes, i.e., the norms that regulate the roles and moves in different types of dialogue. On the other hand, bounded rationality and naturalistic decision making evaluate processes and strategies by their reliability in achieving real world results. Dialogue theory can provide a set of rigorously defined cognitive/social norms, and externalist techniques can gauge the effectiveness of the resulting processes for achieving goals in real world contexts. It provides an empirical basis for the externalist component of dialogue theory.

What Is the Role of Other Persons?

Is thinking an individual or a social enterprise? As shown in Table 4 approaches to critical thinking differ in their answers. According to formal models, reasoning is intensely individual. The aim of critical thinking is to free a person from dogma, superstition, and intellectual laziness, and this task demands a stringent, isolated effort of pure thought. Informal logic, as well as mainstream critical thinking texts, carry on this tradition (in a diluted form) in their emphasis on critical scrutiny of arguments offered by others. Informal logic focuses on *fallacies*, or systematic errors in reasoning, many of which involve susceptibility to unwarranted influence by popular assumptions, unsubstantiated and ill-reasoned opinions published in the newspapers or books, and so on.

Dialogue theory and rhetoric represent a fundamental shift in attitude. Other persons are no longer just dangerous distractions or occasional sources of imperfect information. Dialogue theory *defines* reasoning as part of a social process, in which different individuals verbally interact in a rule-governed way. Reasoning requires two roles, that of proponent and opponent, joined together in a process of challenge and response, in which norms constrain the permissible "moves" for the occupier of each role. Critical thinking by an individual is a mental simulation of such a dialogue in which one individual plays both roles. Unless each of the two essential roles are filled, at least conceptually, there is no critical thought. Thinking presupposes at least the *idea* (if not the actual presence) of another person.

Table 4. Views on the role of other people and their testimony in internalist paradigms.

Field or subfield	**Importance of others in the critical thinking process**	**Conditions for accepting other's beliefs and testimony**
Formal Logic	Reasoning is personal and individual. To be justified, your beliefs must be evident *to you*, either because they are self-evident or you have derived them by self-evident steps. Consideration of opposing views is irrelevant and distracting.	Neither testimony nor judgment of others deserves any credence unless they can be independently established.
Informal logic	The focus remains on how to apply criteria to arguments as products, not to the process of dialogue. On the other hand, dialogue concepts begin to enter the picture, e.g., responding to the objections or anticipated objections of others and adjusting standards of rigor to fit the context.	Supporting a conclusion by appeal to views of others is often a fallacy, e.g., appeal to popularity or authority. On the other hand, testimony may be acceptable as evidence if questions regarding competence, trustworthiness, and experience are answered favorably.
Dialogue theory	Evaluation focuses on argument as an interactive process with a collaborative purpose rather than on products. The specific type of dialogue agreed to by the participants supplies the relevant normative standards for the argument. One type of argument is *critical discussion*, with the purpose of rational resolution of disagreement. This dialogue defines two roles (proponent and opponent) which are essential for critical thought in general.	Testimony can be used as evidence as long as it is *accepted* by the participants in the dialogue, subject to the same process of challenge and defense as any other reasons. The process of defending a position against the challenges of an opponent determines what issues need attention in the reasoning process, i.e., to provide further justification.
Rhetoric	The belief system and values of the audience determine what are the acceptable premises and where lies the burden of proof in argumentation. Evaluation of an argument includes its success in persuading a real audience.	The proponent will base her arguments on the beliefs and values of the current audience, in order to succeed in persuasion, in some cases even if the proponent does not accepts them herself.

This approach has important implications for the treatment of testimonial evidence. Justification in a dialogue context is a matter of successfully defending a belief against challenges. But a defense is unnecessary and indeed is inappropriate if a belief has not been challenged. Beliefs based on statements by other individuals are treated the same as any other claims. On some occasions, acceptance of such testimony may require a response to challenges (e.g., regarding the trustworthiness of the witness). But there is no rigid rule demanding that testimonial evidence always run such a gauntlet. In many contexts, the testimony of others is

accepted almost as automatically as the deliverances of our own senses. Dialogue models provide a framework for this kind of flexibility.

Some rhetoricians and critical thinking theorists go even further in stressing the social nature of thought. They emphasize the social context of reasoning as the source of its validity. Standards of correctness cannot be defined outside the context of a network of shared conventions and assumptions, and are therefore relative to a culture, a context, and to a specific field (e.g., Walters, 1994).

Table 4 summarizes two different but related ways that internalist approaches to critical thinking differ with respect to other people: (i) Others may or may not participate in the process of thinking itself (as a real or hypothetical opponent or audience); and (ii) the beliefs and testimony of others may be acceptable to varying degrees and under varying conditions.

How Can We Decide Among Paradigms?

We started out with an assumption that the purpose of critical thinking is to accept only intellectual products that *should* be accepted, and we have found that there are very different views of what facts make an intellectual product normatively acceptable. In particular, these approaches differ on two key dimensions, both of which regulate the information that is allowed to enter into the evaluation:

- Analytically based models largely exclude empirical findings about actual cognitive or social processes. Empirically based models include such empirical findings.
- Internalist models exclude information not accessible to the awareness of the decision maker. Externalist models include such information.

The question might naturally arise at this point, Why not just settle on the approach that uses the *most information* – i.e., empirical / externalist? After all, what basis is there for excluding information about the environment, external purposes, actual cognitive processes, and cognitive capacity? There is a convincing argument for this position based not simply on inclusiveness, but on the need for a naturalistic approach to critical thinking. Analytical models assume that there is a privileged class of self-evident truths set apart from other beliefs. From the naturalistic point of view, this idea cannot be sustained (Quine, 1994; Kim, 1994).

Unfortunately, the choice is not nearly as simple as this argument implies. These approaches have strengths and weaknesses for different purposes. Each of the two dimensions addressed by the taxonomy in Table 1 involves tradeoffs, as shown in Figure 8. A coherent, plausible reconciliation of these competing views requires mutual adjustment, not just agglomeration, and is not a trivial accomplishment. Here in particular are considerations that such a synthesis will have to address:

Simplicity Versus Comprehensiveness

Analytical criteria tend to be simpler and less ambiguous while empirical criteria tend to be more comprehensive in their coverage of human thinking processes. Empirical models are attractive because they take account of the traits, strategies, and processes that have evolved biologically, culturally, collaboratively, or through the experience of proficient individuals. To serve the purposes of training, critical thinking criteria must be compatible with the cognitive limitations and proclivities of their users. The best way to ensure such compatibility is to model

them after the processes that people already employ. The greater the influence of cognitive theories of actual cognitive processes, the more learnable and understandable the resulting normative theory is likely to be.

On the other hand, empirical approaches may yield a smorgasbord of unrelated criteria reflective of differing standards in different domains. For example, naturalistic decision making studies tend to emphasize specific patterns of cues that decisions makers learn as they become experienced in a particular domain. Informal logic makes a stab at specifying simple criteria (e.g., acceptability of premises, sufficiency and relevance of premises for conclusion, avoidance of specific fallacies). But as noted earlier, simplicity is gained at the expense of ambiguity. None of these criteria is very well-defined, and none can be applied without substantial judgment and even domain-specific knowledge about the substance of the problem.

Analytical approaches promise, at least in principle, greater simplicity. For example, formal logic has a well-defined concept of logical validity; and engineering approaches have well-defined techniques for finding optimal solutions. A *small number* of general principles provide a more economical explanation of what makes an intellectual product acceptable than a large number of specific principles. Training might also be facilitated, since a small set of general criteria might take longer to learn, but can be expected to transfer far more readily to new domains. *Simplicity* thus favors analytically based models (e.g., that propose criteria such as formal validity and maximization of expected utility). *Comprehensiveness* favors empirically based models of normative adequacy, which catalog a large number of different types of argumentation, decision making, and dialogue.

Inputs Versus Outputs

Internal models are attractive because they prescribe rules (e.g., of logical validity, cogency of arguments, appropriateness of moves in a dialogue) that can be applied directly to the information in the possession of the cognizer(s). Inputs are relatively easily obtainable, e.g., by subjective judgment, and no one is blamed for what they do not know. On the other hand, external models are attractive because they address the bottom line, including facts that are unknown to the decision maker but important in their impact on success or failure. External criteria may be more useful because they refer directly to the actual objectives of the tasks in which arguments are generated, and permit blame-free feedback in training. Ultimately what matters is not whether cognitive processes follow any particular set of internal rules, but how reliably they achieve objectives in the real world. Thus, *ease of supplying inputs* favors internalist approaches, while *meaningfulness of output* favors externalist approaches.

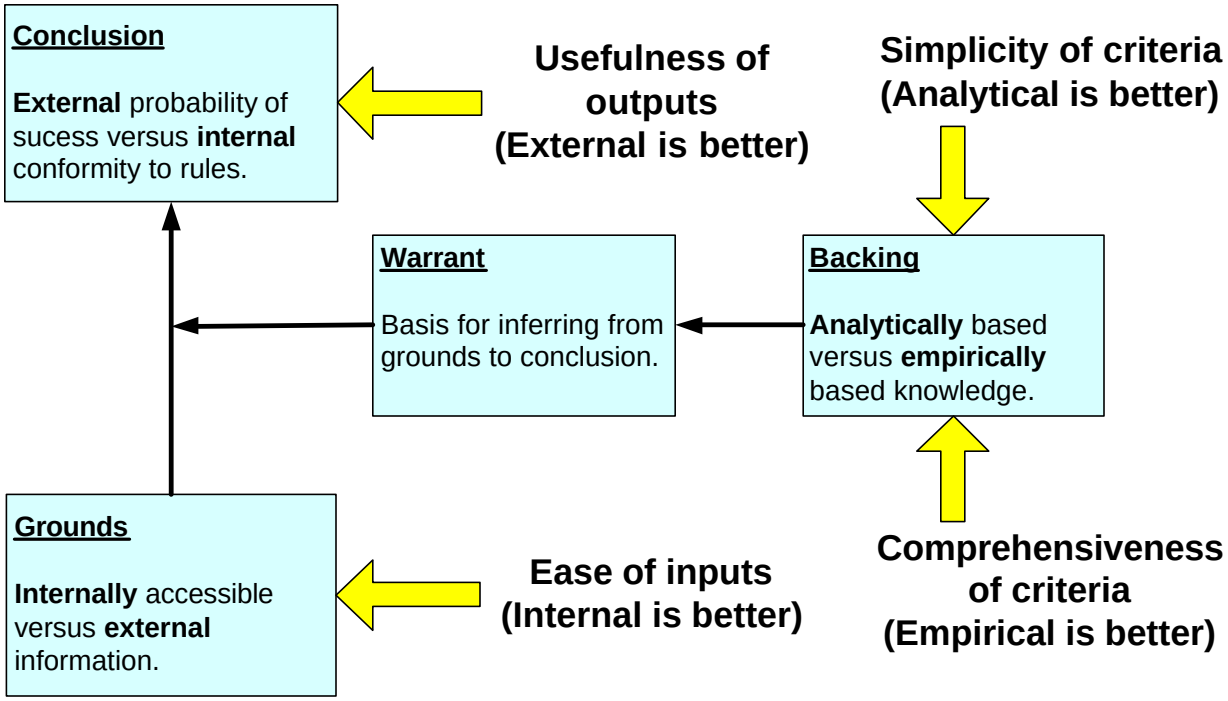


Figure 8. Tradeoffs in integrating different views of critical thinking.

Ultimately, our objective is an integrative framework for critical thinking that addresses these tradeoffs as well as possible. An adequate theory of critical thinking will have the following features:

1. It should offer simple principles which are nonetheless comprehensive, i.e., precise and unambiguous, and tuned to the cognitive processes that people actually employ. It will draw on both normative analysis (for simplicity) and empirical findings (for comprehensiveness). Such a theory will take seriously the way proficient decision makers in fact form beliefs and choose actions, but will also use appropriate analytical tools to capture the normative implications of empirical observations in a reasonably simple and systematic way.
2. An adequate theory of critical thinking will have both internalist and externalist components. It should include both rules that cognizers can apply to regulate their internal processes and external methods for evaluating the effectiveness and efficiency of those rules. The internalist component will specify effective strategies for the conduct of reasoning, while the externalist component will assess the expected benefits and costs of reasoning considered as one among several alternative belief-formation processes.

PART II: A SOLUTION

7. A CRITICAL THINKING THEORY

In this chapter we sketch a theory of critical thinking. More detailed justification and understanding will be provided in the chapters to follow. The full development, application, and testing of the theory will, of course, be a longer term task.

Mental Models, Dialogue, and Reliability

The essence of our theory is that critical thinking skill is exemplified by *asking questions about alternative possibilities in order to achieve some objective.* Asking and answering questions is a skill of *dialogue*. Alternative possibilities are represented by *mental models*. A process of questioning mental models is (or should be) adopted because of its *reliability* for achieving the purposes of the participants within the available time. Thus, the theory of critical thinking draws on and synthesizes research on three separate topics:

1. Theories of reasoning according to which people represent information about a problem or situation by means of mental models of alternative possibilities, evaluate the models in the light of relevant background knowledge, update the models by adding new information as it becomes available, revise models to resolve internal inconsistencies, and draw conclusions by inspecting the surviving possibilities (adapted from Johnson-Laird, 1983; Johnson-Laird & Byrne, 1991).
2. Theories of critical discussion in which a proponent must defend a claim against challenges of various kinds by an opponent or critic (adapted from Rescher, 1977; Walton & Krabbe, 1995; van Eemeren & Grootendorst, 1992; Walton, 1998).
3. Theories of the cognitive mechanisms and processes involved in belief formation and decision making, which vary in their reliability or their association with proficient performance in a domain (adapted from Simon, 1997; Gigerenzer & Selten, 2001; Ericsson & Smith, 1991; Klein et al., 1993; Payne, Bettman, & Johnson, 1993).

Critical thinking, like an onion, has a multi-layered structure (Figure 9). The three aspects of the theory form a spectrum from internal standards and guidance (mental model theory) to external assessment (reliability), with the concept of dialogue forming the crucial link between the two.

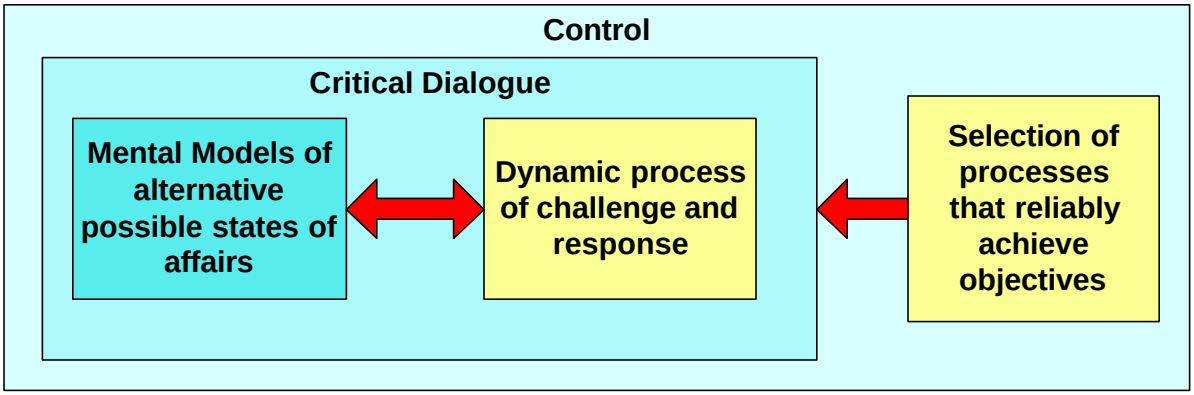


Figure 9. A model of critical thinking with three embedded layers: mental models, critical dialogue, and control based on reliability.

All three of these aspects involve both empirical and normative elements. In particular, each of the three layers is associated with distinctive criteria of performance, which progress from internal to external in their focus:

1. At its innermost core critical thinking involves representation of *alternative possible states of affairs*, or mental models. The key metric of performance at this level is the explanatory coherence of mental models and the coherence between mental models and background beliefs. Errors occur when cognizers overlook alternative possibilities or fail to properly assess the relative plausibility of different mental models, including their comprehensiveness and simplicity as explanations.
2. At the intermediate level, mental models are embedded within a layer of *critical questioning* which motivates the generation and evaluation of possibilities. Such dialogues. Critical questioning may take place within a single mind among different individuals, but is evaluated by reference to norms for conducting the appropriate kinds of critical dialogue. Dialogue types are differentiated by the purposes they serve, by the types of challenges that are permitted to the opponent, and the scope of the permitted responses by the proponent. At this level of analysis, errors occur when cognizers fail to ask or answer questions associated with the appropriate argumentation scheme, use argumentation schemes that obstruct the purpose of the dialogue, or inappropriately shift from one dialogue type to another (Walton, 1998).
3. At the outermost layer, critical thinking is a judgment about the *reliability* of a cognitive process or faculty, hence, the degree of *trust* that should be placed in its outputs. A critical dialogue is only one of various cognitive or social processes that might be utilized alone or in combination to generate beliefs and decisions. Non-deliberative processes, such as pattern recognition, may be more reliable under some conditions and can almost always be used to verify the results of reasoning – just as reasoning is used to check the result of intuition. At this level, errors occur when cognizers use inappropriate or inefficient strategies, and when they terminate a process too soon or continue it too long.

In sum, critical thinking skill is exemplified by *asking and answering critical questions about alternative possible states of affairs, with the intent of achieving the purpose of an on-going activity*.

Mental Models

Johnson-Laird and his colleagues (1983; Johnson-Laird & Byrne, 1991) cite evidence that humans reason in terms of *models* that represent meaning. Comprehending an assertion requires understanding what possible states of affairs are compatible with that assertion (Johnson-Laird & Byrne, 1991). Inference involves at least two processes: (i) combining the meanings of different assertions to determine what states of affairs remain possible if all the assertions are true, and (ii) determining whether the conclusion is true in each surviving possibility. If the results are problematic, an additional process may be employed to (iii) verify that all the relevant possibilities have been considered. The premises of an argument are semantically consistent if there is a possible state of affairs in which they are all true. An argument is *semantically valid* if there is no state of affairs in which the premises are true and the conclusion is false.[43]

According to Johnson-Laird (l983), what distinguishes a mental model from other representations is a close structural isomorphism between the model and the state of affairs it represents. Numerous related facts can be combined with one another in the same representation, and the implications of the combined facts can then simply be recognized or read off the representation without the need for logical deduction or for the separate explicit statement of each implied fact that is characteristic of logical or probabilistic models. The simplest example is a map, which uses symbols and spatial relations to represent objects and their spatial relations. If we learn that enemy army A is west of division B, and division B is west of army C, we know that army A is west of army C; the relation "to the left of" among the symbols for A, B, and C on the map preserves this same transitivity. The relations represented by mental models need not be spatial, but may (for example) be temporal, causal, or conceptual; the relations used in the representation may be, but need not be, the same type as the relations being represented (although they must be isomorphic). The concept of a mental model contrasts with representations such as semantic networks, predicate calculus, or, indeed, ordinary English grammar in which numerous formal devices play no direct symbolic role. A similar concept can be found in Shepard's (1975) discussion of first- and second-order representational isomorphism.

According to mental model theory, the difficulty of an inference increases with the number of alternative possibilities that must be constructed to solve the problem, as well as the complexity and familiarity of the possibilities. Errors may occur for several reasons: The number of possibilities exceeds capacity limitations of working memory (Johnson-Laird & Byrne, 1991, p. 39); there is a tendency to represent only possibilities that are true given an assertion, and only components of assertions that are true within a possibility – and thus to neglect possibilities consisting of false components; a prior tendency to believe that the conclusion is correct causes

[43] In addition to work on the psychology of reasoning by Johnson-Laird and his colleagues, logicians have also developed formal systems of syntactic rules based more closely on semantics (e.g., Hintikka, 1999). The syntactic rules in such systems are explicitly designed to guide the reasoner in the search for a model that is consistent with the premises and in which the conclusion is false. If no such counter-model can be found, the argument is valid. The close relationship between syntax and semantics makes the soundness of such logical rules virtually transparent. The process of proof itself is essentially a critical thinking strategy, i.e., looking for ways the conclusion could be false.

the reasoner to cut short the exploration of alternatives. Because of such limitations and biases, people are liable sometimes to accept a conclusion even though there is a possible state of affairs in which it is false.

The representation of a particular possible state of affairs is called a *mental model*. Typically, a simple statement (e.g., *the enemy has developed long range artillery*) is compatible with many different states of affairs (how they developed the artillery, when, where, how long its range is, etc.), but to conserve processing capacity, people typically select a single, representative mental model to depict its meaning (Johnson-Laird & Byrne, 1991, p. 170). They elaborate the level of detail of the representation and/or expand it to include other possibilities only when forced to do so by new information (or critical thinking challenges). Thus, implicit assumptions, the possibility of retraction, and the need for verification by critical thinking are essential elements of reasoning at ground level, in comprehending the meaning of the premises.

When a sentence contains logical connectives, such as *and*, *or*, *if-then*, *all*, and *some*, people use knowledge of their meaning to construct the appropriate set of mental models. Deductive reasoning is the process of combining assertions of this kind into a single set of mental models and extracting conclusions from the surviving models. Suppose that the first premise in a deductive argument is *p or q or both* (e.g., *the enemy will attack either through the northern pass or through the southern pass or both*). The premise contains two elementary propositions, *p* and *q*, and there are four possible combinations of their truth and falsity. One possibility, in which neither p nor q is the case, is false if the assertion is true. Thus, the meaning of the premise may be represented in terms of the following three possibilities, or mental models:[44]

1	p	not-q
2	not-p	q
3	p	q

When multiple assertions are made, their meanings are also interpreted in terms of possible states of affairs. According to mental model theory, the heart of the *inference* process is a normal *comprehension* process, in which the cognizer integrates the representations of several statements in such a way that the surviving mental models are compatible with all of the statements. To see what, if anything, follows from the premises, she determines what if any new facts are true no matter what, i.e., across all the surviving mental models.[45] In our example,

---

[44] According to Johnson-Laird & Byrne (1993), people usually do not explicitly represent the false components of an assertion. Thus a more economical format will be used:

1	p	
2		q
3	p	q

People make the false elements explicit only when necessary for the inference.

[45] Alternatively, to verify a *proposed* conclusion, she determines whether it is in fact true in all the remaining mental models.

suppose the second premise is *not-q* (e.g., we learn or infer that *the enemy will not attack through the southern pass*). Then mental models 2 and 3, in which q is true, are ruled out, and the only remaining possibility is:

1	p	not-q

The cognizer now examines the surviving model to see if it can be summarized parsimoniously and non-redundantly, i.e., in a way that does not simply repeat a premise. If there is a statement that satisfies these conditions, it is the *conclusion* of the inference, which in this case is *q* (e.g., *the enemy will attack through the northern pass*).

Mental model theory has been tested by verifying its predictions about the degree of difficulty of such inferences and the nature of the errors that occur (e.g., Johnson-Laird & Byrne, 1991, pp. 52-56). Suppose we assert a conditional *If p then r* (e.g., *If the enemy intends to attack in the north, they will locate artillery in the north*). This conditional is compatible with all states of affairs except the one in which *p* is true and *r* is false. However, to conserve working memory, people tend to ignore the possibilities in which the antecedent *p* is false, and thus represent conditionals by a single mental model:

1	[p]	r
	…	

The "…" is an annotation or reminder that all the models consistent with the premises are not represented, i.e., that some possibilities remain implicit. The square brackets around the p restrict what these implicit models can contain. They indicate that p is exhaustively represented with respect to r. That is, since p is explicitly shown only in association with r, p cannot occur in conjunction with not-r in any other models. The implicit possibilities will be fleshed out only if necessary.

If the second premise is *p*, (e.g., *the enemy intends to attack in the north*) the conclusion *r* (*they will place artillery in the north*) follows directly from model 1. The theory thus predicts that this type of inference (called *modus ponens*) will be easy. But suppose the second premise is *not-r* (e.g., *the enemy has not placed artillery in the north*). This premise is inconsistent with the abbreviated representation of the conditional premise *If p then r*, because it falsifies model 1. The default assumption (*p and r*) has been *defeated*. If the premises are in fact consistent with one another, there must be additional possibilities that were not part of the cognizer's initial representation. To get the correct answer, therefore, the cognizer must search for these additional mental models and add them to the representation. She can then see what conclusion, if any, follows. The conditional is exhaustively represented by three mental models:

1	p	r
2	not- p	r
3	not p	not-r

where (2) and (3) flesh out the previously implicit model (the three dots). The premise *not-r* rules out mental models (1) and (2), leaving a single possibility (3), in which both *not-r* and *not-p* are true. Since *not-r* was a premise, the correct (non-redundant) conclusion is *not-p* (i.e., *the enemy will not attack through the northern pass*). The theory predicts that inferences of this kind

(*modus tollens*) will be relatively difficult compared to the previous problem (Johnson-Laird & Byrne, 1993, pp. 49, 53).

In complex inferences, there may be a stage in which the cognizer *verifies* her conclusion. To do so, she checks to see if she has in fact generated all the relevant alternatives that are compatible with the premises and with her background knowledge. She then examines these models to ensure that the conclusion is true in each. If she finds models in which the conclusion is not true, she has several options. First, she may drop the conclusion and look for a new one. She examines the set of mental models to see if there is any statement that summarizes them in a novel, parsimonious way. Secondly, the cognizer may summarize by stating that the original conclusion is *possible* although not necessarily true, because it is true in at least *one* model that is consistent with the premises. Thirdly, she may look at the *proportion* of mental models in which the conclusion is true and summarize the result probabilistically. For example, if there are 10 equally probable mental models and the conclusion is true in two of them, its probability may be estimated at 20% (Johnson-Laird, Legrenzi, Girotto, Legrenzi, & Caverni, 1999). In real-life situations (which usually do not involve deductive logic), she may also have the option of trying to collect or remember more information or evidence, to further narrow down the set of mental models.

Situation Models, Causality, and Background Beliefs

It will be handy to introduce some terminology of our own that is especially suited to critical thinking in real-time military domains. We define the cognizer's *explicit situation model* as the set of mental models that represent, at a given time, the cognizer's understanding of a dynamic real-time situation. On this definition, an explicit situation model has two dimensions. The numbered rows represent distinct possible accounts or assessments of the current situation, i.e., mental models. The columns represent the variables with respect to which the mental models are discriminated from one another. (Often these are propositional variables that take the values *true* or *false*.)[46] An explicit situation model tells us what the cognizer consciously believes to be the case. But it also tells us what aspects of the situation the cognizer regards as *relevant*, as reflected in the variables used to distinguish possibilities. And it tells us which relevant aspects of the situation are *uncertain*, i.e., those variables for which multiple values are instantiated in different rows. From this perspective, situation understanding is more than knowing facts. It includes knowing which facts are important in the current situation, and which of these facts are uncertain. All three are necessary inputs to deliberative strategies for reducing uncertainty.[47]

A situation model in which the variables are causally related to one another is a *causal mental model*. A causal mental model in which the variables refer to a human action (and its causes and effects) is a *story* (cf., Hastie, 1993). A causal mental model in which one or more of the variables is not observed but is inferred in order to explain other variables, is called a

---

[46] If no other structure is represented, the situation model is analogous to a table in a relational database, where the rows are "cases" and the columns are "fields." Note that this is not intended to be only one way to represent a situation model; additional structure (e.g., causal or semantic) can also be shown..

[47] SAGAT (Endsley, 1997), a well-known tool for measuring situation awareness, involves stopping a scenario at randomly selected points and querying the subject regarding the values of randomly selected variables. The variables to be queried are selected from the same pool for each scenario intervention. This method overlooks the importance of knowing *which* variables are worth attending to at different points in the scenario. It also overlooks the importance of recognizing uncertainty regarding such variables.

situation theory. The general knowledge or schemas that underlie the construction of a situation theory is a *background theory*. For example, military officers build stories about enemy intent (which is not directly observed), in order to explain and predict future enemy actions (Cohen, Thompson, Adelman, Bresnick, Shastri, & Riedel, 2000a). Underlying the construction of such a situation theory / story is a background theory containing general causal knowledge about the way people behave and more specific causal knowledge about enemy motivation and capabilities. The causal structure of this background knowledge influences the choice of variables to use in mental models. These variables, in turn, determine the *questions* that are asked in critical thinking.

Each dimension of the explicit situation model is subject to change in the course of critical thinking. First, new variables may be introduced as questions are asked and answered. These new variables in turn will result either in the addition of new mental models (increasing uncertainty) or in the elimination of some of the currently existing possibilities (decreasing uncertainty). Critical thinking therefore alters both the number of variables and the number of mental models under consideration over a series of "moves" which are ideally part of a larger strategy. The ultimate intent of the strategy is to reduce uncertainty in situation understanding, i.e., eliminate a sufficient number of mental models to support a conclusion, which in turn assists in the achievement of a purpose. Part of critical thinking skill is the ability to select effective and efficient question-and-answer strategies for searching in a space of mental models.

As noted, the explicit situation model is not all there is to situation awareness. A significant component of an individual's knowledge is embedded in a rapid, parallel processing system that includes the spread of activation over a network of beliefs (e.g., Cohen, Thompson, Adelman, Bresnick, Shastri, & Riedel, 2000c). Such a system is necessary as the source of background beliefs by means of which people are able to construct and evaluate explicit mental models. Critical thinking about mental models is a *deliberative* method of improving situation awareness in order to achieve a goal. In doing so, it explores and exploits a large background system of beliefs, most of which remains implicit. Dialogue theory is a source of insight and a tool for modeling the strategies by means of which such a critical thinking process is conducted.

Dialogues

The field of informal logic has lacked a unifying theory that successfully accounts for different types of arguments and the errors to which they are subject (Walton, 1998, p. 7). A promising approach which is drawing increasing attention is the interpretation of argument as a component of *dialogue*. As Johnson (1996) says, "an argument understood as *product* – a set of propositions with certain characteristics – cannot be properly understood except against the background of the process which produced it – the process of argumentation." Dialogue theorists attempt to describe argumentation by means of rigorous, idealized models of interactive exchanges. Such models specify the purposes of different types of dialogue, the roles that are played within the dialogue, rules for each player, and rules for determining who wins. Actual discussions can be analyzed and evaluated in terms of how closely they approximate the appropriate paradigm (Walton & Krabbe, 1995, pp. 174-177).

Prominent contributions to dialogue theory have been made by researchers in Amsterdam (viz., van Eemeren and Grootendorst) and in Canada (Walton). The pragma-dialectical theory proposed by van Eemeren and Grootendorst (1992, 1994) closely interweaves normative and descriptive elements. An ideal of critical rationality in dialogue is developed, and at the same

time actual processes of argumentative discourse are studied empirically. Actual argumentative discourse is reconstructed from the perspective of the ideal of *critical discussion*. This permits the discovery of practical problems or errors experienced in argumentative discourse, and forms the basis for development of appropriate methods in education (van Eemeren, Grootendorst, & Henkemans, 1996). The source of the norms is an ideal of actual human discourse, rather than a formal axiomatic system (as in logic or probability theory). According to Walton,

> A dialogue is a conventionalized framework of goal-directed activity in which two participants interact verbally by taking turns to perform speech acts. Typically, these speech acts are questions and replies to questions. The various speech acts are linked together in a sequence that has a purpose and direction as the dialogue proceeds. The purpose is determined by the goal of the dialogue as a recognized type of social activity. (Walton, 1998, p. 98)
>
> A dialogue is a goal-directed, collaborative conversational exchange, of various types, between two parties. … fallacy is defined as an argument or a move in argument that interferes with the goal of a dialogue of which it is supposed to be a part…. (1996b)

Among the central themes in recent work on dialogue theory are the following:

- Classification of multiple types of dialogue (Walton, 1998)
- Analysis of critical discussions (or persuasion dialogues) into stages (van Eemeren & Grootendorst, 1992)
- Identification of top-level principles for each stage of a critical discussion (van Eemeren & Grootendorst, 1992)[48]
- Understanding fallacies as violations of the top-level principles of a dialogue stage (van Eemeren & Grootendorst, 1992)
- Developing formalized models for questions and replies in the argumentation stage of a critical discussion (Hamblin, 1970; Walton & Krabbe, 1995; Rescher, 1977)
- A theory of argumentation schemes for determining where the presumption of truth lies and what critical questions are appropriate (Walton, 1996a)
- Understanding fallacies as either misuse of an argumentation scheme or an illicit shift between types of dialogue (Walton, 1995)

We will briefly touch on these in turn.

Dialogue Types

A variety of different dialogue types have been identified by Walton (1998; Walton & Krabbe, 1995, p. 66), including *persuasion, deliberation, negotiation, information seeking,* and *inquiry*. They differ, according to Walton and Krabbe, in their main goals and in the initial situations that they address. We present these in a scheme that is a modified version of one

[48] van Eemeren and Grootendorst call these *rules*. We will refer to them as *principles* to distinguish them from the more fine-grained rules that govern speech acts in formalized models of dialogue.

suggested by Walton and Krabbe: (i) A dialogue may aim at changes in belief, in action, or in affect. (ii) A dialogue may or may not involve conflict between the parties; if it does not, it may be a symmetric non-conflictual condition or an asymmetric non-conflictual condition. Walton and Krabbe mention the following general dialogue types, which fit within our scheme as follows:[49]

Aims at change in beliefs

- Persuasion (or critical discussion) – Resolve *conflicting* points of view
- Inquiry – Expand knowledge that all parties need (*symmetric*)
- Information Seeking – Spread knowledge when one party is ignorant (*asymmetric*)

Aims at change in actions

- Negotiation – Resolve *conflicting* interests
- Deliberation – Reach a shared decision for all parties (*symmetric*)

Aims at change in affect

- Eristic (Quarrel) – Reach an accommodation in a *conflictual* relationship

Crossing the two dimensions (3 x 3) yields nine cells. Thus, our taxonomic framework suggests three additional types of dialogues to fill in the missing cells. Table 5 shows the full matrix, with the suggested new categories in italics. The new categories seem especially relevant for an extension of dialogue theory to the analysis of leadership and critical thinking in teams.[50]

[49] Walton and Krabbe's classification (p. 80) is less systematic. For example, instead of action-belief-affect, their goal dimension includes the following three items: *Stable agreement/resolution, Practical settlement/decision (not) to act, and Reaching a (provisional) accommodation*. The problem is that resolution-settlement-accommodation are (i) ambiguous terms, and (ii) not clearly coordinated with one another, i.e., mutually exclusive and at the same level in a hierarchy. Walton and Krabbe's other dimension (the initial situation) includes: *conflict, open problem, and unsatisfactory spread of information*. Again, the terms are not precise and not clearly coordinated. An additional drawback is that Walton & Krabbe's two dimensions do not cross with one another to produce a fully populated matrix of dialogue types.

[50] A natural elaboration of the taxonomy would involve distinguishing symmetrical and asymmetrical subcategories within conflictual dialogues. There is already a distinction between persuasion dialogues in which each party defends a thesis against the other (symmetric) and persuasion dialogues in which one party is the proponent and the other is the opponent (asymmetric). Similarly, in dialogues centered on actions, negotiations involve a two-sided determination of action, but regulatory or supervisory arrangements are one-sided. Finally, quarrels can also be either one or two sided.

Table 5. Dialogue types classified by main goal (rows) and initial situation (columns).

	Conflictual	**Non-conflictual**	
		Symmetric	**Asymmetric**
Modify Belief	Persuasion	Inquiry	Information-seeking
Modify Action	Negotiation	Deliberation	*Guidance / advice / instruction -seeking*
Modify Affect	Quarrel	*Team forming, mutual identification, bonding*	*Exhortation, seduction, conversion*

According to dialogue theories, participants cooperate to choose the type of dialogue that is best for the purpose and context (van Eemeren & Grootendorst, 1992). Hence, they must make implicit or explicit judgments about the relative reliability of different dialogues as methods for achieving their goals. They must also reflectively monitor adherence to the norms that govern the relevant type of dialogue (Jackson, 1989; Johnson, 2000). Some dialogue types are profitably embedded within others (Walton & Krabbe, 1995, p. 73). For example, a persuasion dialogue may be suspended in order to settle a matter of fact by means of an information seeking dialogue or inquiry. A negotiation may benefit from an inquiry, persuasion dialogue, or information seeking dialogue to resolve disputes about facts. However, as Walton and Krabbe point out, some dialogue shifts are not beneficial, and in fact correspond to fallacies. The *fallacy of bargaining* occurs when a persuasion dialogue becomes a negotiation, i.e., when the parties to a disagreement of beliefs attempt to purchase one another's agreement. More generally, it is a fallacy to believe something because it is in one's practical interests to believe it (e.g., a cigarette executive chooses to believe that cigarettes are healthy).

Stages, Principles, and Fallacies

Principles governing the possible actions of each dialogue participant are a function of the type of dialogue, the role being played by that participant, the stage of the dialogue, and the dialogue history (i.e., previous statements of the participants). According to van Eemeren and Grootendorst (1992, pp. 34-37), a *critical discussion* is a type of dialogue used for the resolution of a difference of opinion between a proponent and an opponent. Resolution is not a matter of negotiation or of ignoring differences in views. Resolution can occur in only two ways: The proponent *persuades* an opponent to retract doubt concerning the proponent's position because she has been *convinced* by the proponent's reasons, or conversely the proponent relinquishes her position because it has not withstood the opponent's challenges. A critical discussion is a special case of Walton's (1998) *persuasion* dialogue.[51] A critical discussion has four stages, each of which is associated with top-level principles. The principles are based on Grice's norms for

[51] In other versions, such as Rescher (1977), there are three participants or roles: a *proponent*, an *opponent*, and a *judge*.

cooperative conversation (Grice, 1989). A fallacy in reasoning, according to van Eemeren and Grootendorst (1992) is a violation of one of those principles.[52]

Stage 1. In the *confrontation* stage, a difference of opinion is acknowledged. For example, the proponent expresses a standpoint with or without reasons. The opponent asks questions to clarify or flesh out what the proponent meant to say and indicates disagreement or expresses doubt.

The most important principle for this stage is that the parties "must not prevent each other from advancing or casting doubt on standpoints" (van Eemeren & Grootendorst, 1992, p. 208). Fallacies that involve violation of this rule include: threats, personal attacks on competence or motives, appeals to sympathy of other party, or declaring standpoint immune to criticism.

Stage 2. In the *opening* stage (van Eemeren & Grootendorst, 1992, p. 41), the parties "agree" (perhaps implicitly) on the type of discussion they will have and the discussion rules. For example, in a critical discussion, they agree that one party will take the role of proponent and the other will take the role of the opponent. The major difference between the proponent and opponent in a critical dialogue is the global *burden of proof*. In a *simple* critical discussion, it is up to the proponent to create a positive case for her standpoint. The opponent merely has to create doubt, not to positively defend any thesis. A *compound* dialogue, on the other hand, is symmetric. The two parties defend contrary theses, and each participant plays opponent to the other. The difference between simple and compound critical discussions turns out to be a highly significant distinction in terms of the moves available to the parties and the depth and richness of the reasoning.

The top-level principle for the opening stage concerns the burden of proof: Whoever advances a standpoint is obliged to *defend* it if asked to do so by the other (p. 208). Fallacies that violate this rule include: Evasion by giving personal guarantee that thesis is correct, evasion by declaring the standpoint obvious, and turning the tables and making the challenger prove that the standpoint is wrong (rather than demanding that the proponent show it is correct). In a compound dialogue, where both parties are proponents, it is a fallacy to require only one party to defend her standpoint.

Stage 3. The central stage of a critical discussion is *argumentation*. During this stage, the proponent advances reasons to believe the standpoint, the opponent expresses doubt, the proponent defends, and so on.

van Eemeren and Grootendorst spell out a number of top-level principles for the argumentation stage, each with associated fallacies. One principle is that a standpoint may be defended only by advancing arguments relating to that standpoint. This rule ensures *relevance*. It rules out such fallacies as appeals to the emotions of the audience, appeals to one's own authority, use of false modesty, and arguing for something different than the real issue under discussion. A second principle for the argumentation stage is that a person can be held to the premises he leaves implicit. Violations of this principle include denying a key premise just because it wasn't expressed, and creation of a straw man by exaggerating the protagonist's unexpressed premises. A third principle is that parties can be held to all and only the premises

---

[52] van Eemeren and Grootendorst refer to these as *rules*. We call them *principles* because of their general nature and to distinguish them from more specific rules in formalized models at the speech act level.

Walton (1995) gives a more complex analysis of fallacies in terms of argumentation schemes.

agreed to as the accepted starting point. Violations include: begging the question (i.e., taking what is supposed to be the conclusion as a premise), hiding presuppositions in loaded questions (*Have you stopped beating your wife?*), challenging the agreed starting point or trying to add to it, or presenting a new premise as self-evident.

Another principle is a kind of catch-all, that a standpoint is conclusively defended only if the defense takes place by means of correct application of a commonly accepted "scheme of argumentation." This means avoiding commonly recognized errors in reasoning such as: appealing to bad or good consequences of accepting a standpoint, appealing to majority opinion or authority, using misleading analogies, inferring causality from juxtaposition of events, generalizations from inadequate sample, and false appeal to a slippery slope. A final principle is similar but refers to avoidance of "logically invalid" reasoning, such as confusing necessary and sufficient conditions or part and whole.

Stage 4. In the *concluding* stage of a critical discussion, the dispute ends either because the proponent withdraws her thesis or because the opponent withdraws her doubt. A top-level principle for this stage is that a failed defense must lead the proponent to withdraw her standpoint, and a successful defense must result in the opponent withdrawing her doubt. It would be a fallacy in this stage to conclude that a standpoint is certainly false just because it was not successfully defended, or that a standpoint is certainly true just because it was not successfully challenged.

Formal Dialogue Rules

Formalized models of the argumentation stage of a persuasion dialogue have been developed to analyze different ways that such discussions can be conducted. In these models the proponent and opponent take turns speaking, and the rules spell out the types of assertions that are permitted to each side. For example, Walton and Krabbe (1995, p. 149), define two quite different types of persuasion dialogue: rigorous and permissive. A rigorous persuasion dialogue (RPD) is asymmetric, viz., only one party plays the role of proponent, and the role of the opponent is tightly constrained. In addition, it does not permit retraction of assertions and provides no role for implicit, background beliefs.

A permissive persuasive dialogue (PPD), on the other hand, is symmetrical, allows free questioning and challenge, permits retraction, and bases permissible moves in part on implicit background beliefs of the parties, which Walton and Krabbe call "dark-side commitments" because they are not visible. A party's turn may include multiple instances of the following speech acts:

Change own commitments

- Assert
- Retract assertion
- Concede
- Retract concession

Try to influence commitments of other party

- Challenge
- Request concession (yes/no question)

- Request resolution (of incoherence within a model)
- Request reconsideration (i.e., resolution of incoherence with background beliefs)

Although the global burden of proof is static (and rests upon the proponent), as each side provides arguments or challenges, the local burden of proof switches back and forth (Rescher, 1977, p. 27). That is, whenever either side advances an argument, it stands until explicitly rebutted by the other side. The following are paraphrases of some of the rules proposed by Walton and Krabbe (1995, pp. 150-152) for the two parties in a permissive persuasion dialogue. We state them in terms of Prop and Op to indicate which role a given rule targets, but the symmetry of the dialogue means that each role (and rule) applies to both players. It is convenient to classify them in terms of one party's efforts to influence the beliefs of the other party:

Rules for challenges

- If Prop asserts A, Op must either concede A or challenge A. (As a special case: If Prop defends A, for each element of Prop's argument that is not already conceded or challenged, Op must concede or challenge it.)
- Op can challenge A only if Prop asserted A, not merely conceded A.
- If Op challenges A, Prop must either defend A or retract A
- Prop may defend A only if Op has challenged A.

Rules for requesting concessions

- If Prop asks for concession of A, Op must either concede A or not concede A.
- Op may concede A only if Prop has asserted A or asked for a concession of A.

Rules for requesting resolution of incoherence within a model

- If Op has conceded both A and B, and they are inconsistent, Prop may demand that Op resolve A and B.
- If Prop demands that Op resolve A and B, Op must retract concession of either A or B. Op must also retract other commitments leading to that commitment.

Rules for requesting resolution of incoherence with background beliefs

- If Op refuses to concede A or challenges A, then Prop can request reconsideration of A.
- If Prop requests reconsideration of A, and A is a dark-side commitment of Op (i.e., is among Op's background beliefs), Op must concede A.

The proponent incurs an obligation to defend or modify her standpoint at each move, and the opponent incurs an obligation to accept or reject the proponent's assertions at each move. Each assertion must support the goal of the dialogue type they have selected, e.g., to resolve the difference of opinion, and the parties must not to shift dialogue types without mutual agreement. Each party also has an obligation to use words clearly and consistently. Such rules keep the dialogue moving, ensure relevance, and increase the chance of resolution.

Argumentation Schemes, Burden of Proof, and Fallacies

Walton (1996a) supplements general dialogue rules with more specific ones that are based on particular argumentation schemes. Argumentation schemes, according to Walton, are normatively binding kinds of reasoning, i.e., appropriate moves and countermoves in a dialogue. Walton regards fallacies as legitimate argumentation schemes that have been applied inappropriately or employed incorrectly Among the argumentation schemes described by Walton are the following:

Case-based reasoning

- Argument from example
- Argument from analogy

Causal reasoning

- From effect to cause
- From cause to effect
- From correlation to cause

Rule-based reasoning

- From established rule
- Argument for exceptional case
- Argument from precedent
- Argument from pity

Practical reasoning

- Argument from good or bad consequences
- Argument from waste (sunk costs)
- Argument using threat

Gradualistic reasoning

- Device of stages
- Causal slippery slope
- Precedent slippery slope
- Verbal slippery slope

Position to know reasoning

- Position to know
- Testimony
- Expert opinion
- Ignorance

Source indicators reasoning

- Argument based on character of source
- Argument from bias of source
- Argument based on opinion of large majority

Each argumentation scheme is associated with (i) typical premises and a presumptive (i.e., default) conclusion, and (ii) a set of critical questions that may be used by the opponent to challenge the conclusion. The key feature of argumentation schemes is that they reverse the usual burden of proof. A proponent who uses an argument scheme correctly has presumption on her side. That is, the proponent is not obligated to defend the default conclusion in response to generic challenges, as long as the conclusion was arrived at by correct use of the argument scheme. The burden is on the opponent to issue *specific* challenges based on the critical questions associated with that scheme. Asking an appropriate critical question shifts the weight of presumption to the opponent, and satisfactory answers by the proponent shift the weight of presumption back to the opponent. Argumentation schemes allow argument to move forward even when there is insufficient evidence at a particular point in time; hence, they are especially useful when timely action is necessary. Misuse of such schemes, however, is associated with fallacies.

Among the traditionally recognized fallacies is one called *argument from authority* (*ad verecundiam*). In committing this fallacy, the proponent shields her thesis from attack by accusing the critic of showing disrespect for some authoritative source, such as an expert. The problem, as Walton points out, is that it is not always a fallacy to appeal to an expert source. Walton introduces argumentation schemes as a way of distinguishing fallacious uses of a particular form of argument from legitimate, non-fallacious uses – even when the latter are incomplete and non-conclusive. In the argumentation scheme for appealing to *expert opinion* (Walton, 1996a, p. 64-67), there are several premises: *x is an expert in domain D*, *x asserts that A is known to be true*, and *A falls within the domain of expertise D*. These lead to a default conclusion: *Therefore, A may be taken to be true*. Thus, a proponent is entitled to claim that A is true because x, who is an expert on such matters, said so. The opponent cannot simply respond with nothing more than a general challenge (e.g., *Why do you believe A?)*. That standard has already been met. The opponent must work a bit harder and raise *specific* critical questions in order to shift the burden of proof to the proponent, such as: *Is x a genuine expert in D? Did x really assert A? Is A relevant to domain D? Is A consistent with what other experts in D say? Is A consistent with known evidence in D?* The question, *Did x really assert A?* in turn might be challenged with specific sub-questions: *Is the opinion directly quoted or can it be checked? Does it look as if important information or qualifications might have been left out? If more than one source is cited, are they quoted separately? Is the statement clear, e.g., are technical terms explained?*

A variety of different fallacies can be involved in appeals to expert opinion. Such an appeal may be used to squelch critical questioning on the grounds that one party has no right to question the expert's judgment. This is a fallacy that hinges on evasion of the burden of proof. Other errors involve failure to meet the conditions of the argumentation scheme. For example, an expert in one field is sometimes treated as an expert in other fields via a "halo" effect. The expert may not be named or the expert's opinion may be misstated. When the conditions of the argumentation scheme are met and the argumentation scheme is appropriate for achieving the

goals of the relevant type of dialogue the appeal to expert opinion is a legitimate form of argument.

Another example of an argumentation scheme is reasoning by analogy. The premises and conclusion are: *Generally, case C1 is similar to case C2. A is true (false) in C1. Therefore, A is true in C2.* This reasoning can be challenged by such critical questions as: *Is A really true in C1? Are C1 and C2 similar in the respects cited? Are there important differences between C1 and C2? Is there some other case C3 that is also similar to C1, but in which A is false?* These critical questions help avoid false or misleading analogies.

Dialogue theory provides a deeper analysis of fallacies than the usual description in informal logic, in terms of surface features. For example, one fallacy, which is the converse of appeal to expertise, attacks a thesis by attacking the person who proposes it (*ad hominem*). A simple example of a rule of discourse emerging from dialogue theory is the following: "Parties must not prevent each other from advancing standpoints or from casting doubt on standpoints" (van Eemeren & Grootendorst, 1992, p. 208). *Ad hominem* fallacies sometimes involve violation of this rule by the opponent, just as appeals to expertise sometimes involve violations of the same rule by the proponent. According to dialogue theory, when personal attacks are intended to prevent an opponent's views from being fairly considered, the violation of the rule of cooperation governing the dialogue is what makes this an error, not its surface features ("attacking the person"). Other fallacies (e.g., argument by appeal to pity or threats of force) that are different from *ad hominem* and appeal to expertise on the surface appear to involve violation of the same dialogue principle and thus are the same error at a deeper level. Conversely, in other contexts, impugning the character of a person may be appropriate, e.g., if the credibility of a person's testimony must be relied on in drawing a conclusion, as in the *testimony* argumentation scheme. Understanding errors in terms of dialogue rules and argumentation schemes is a step toward a more nuanced and more accurate assessment of their normative status.[53]

Critical Discussion and Critical Thinking

The critical discussion (or *persuasion* dialogue) provides a promising framework for both understanding and training critical thinking. The primary reason for its usefulness is the functional similarity between rationally persuading another individual to accept or reject a position, and rationally determining for oneself whether a position is acceptable or not. The same parallel between inner and outer applies to some of the other dialogue types as well, such as deliberation, inquiry, and negotiation, which are other symmetric dialogues. There is a parallel to negotiation, for example, when an individual debates with herself about tradeoffs among her own competing objectives. A dialogue externalizes necessary functions that must take place within an individual cognizer. Thinking may be fruitfully studied as a form of internal dialogue in which a single individual takes on distinct dialectical roles (Walton, & Krabbe, 1995, p. 26).

Another reason for focusing on dialogue as a model of thinking is that thinking and dialogue share a developmental history, not just a functional similarity. A variety of

---

[53] This framework raises some issues, however, which we will address later. Are argumentation schemes entirely conventional or can they be accounted for on some non-arbitrary basis, for example, as causal models of the world? What is the basis for the distinction between premises (which must be addressed) and critical questions (which are optional)? Is the distinction hard and fast, or is it variable? Integration with mental model theory will help answer these questions.

developmental psychologists (starting perhaps with Vygotsky) have proposed that thinking first develops in each individual as internalized speech and that we learn to reflect on and evaluate our own thoughts by responding to the thoughts of others (Bogden, 2000). As noted by Rieke and Sillars (1997),

> …research suggests that critical thinking is really a mini-debate that you carry on with yourself. What is often mistaken for private thought is more likely an "internalized conversation (Mead), an "internal dialogue" (Mukarovsky), or an "imagined interaction" (Gotcher and Honeycutt).

A final reason for interest in dialogue theory is more direct. Much critical thinking takes place in a team or group context, in which dialogue plays a *literal* role in decision making. The road to improved critical thinking in both an individual and a team context may lead through training in improved skills and habits for critical dialogue. Other dialogue types may also have a direct bearing on performance in teams and groups.

Mental Models and Dialogues

Commitment in Dialogues and Mental Models

The argumentation stage of a critical dialogue is a process of constructing, evaluating, and modifying mental models. At a more specific level, dialogue theory links up with mental model theory via its concept of a *commitment store* (Hamblin, 1970; Rescher, 1977; Walton & Krabbe, 1995). According to Hamblin (p. 257), " a speaker who is obliged to maintain consistency needs to keep a store of statements representing his previous commitments, and require of each new statement he makes that it may be added without inconsistency to this store…" Walton and Krabbe (1995) distinguish two kinds of explicit commitment stores: strong commitments based on assertions, which the party is obligated to defend, and weak commitments based merely on concessions, which the party is not obligated to defend. Rules for permissible moves in the argumentation stage of a dialogue refer to the current status of these commitment stores, and specify how each move changes their contents. For example, consider the first two rules mentioned earlier:

The opponent can challenge any strong commitment of the proponent as long as it is not in the opponent's own commitment store. If the opponent challenges a strong commitment of the proponent, the proponent must either defend it by supplying reasons or retract it. When the opponent does not immediately challenge an assertion by the proponent, the listener has conceded it at least temporarily, and it goes into the listener's weak commitment store. The opponent is not obligated to defend her concession, but must allow the proponent to use it in argumentation at least for the time being. The opponent can retract the concession at any time by challenging it, as long as it is still in the proponent's strong commitment store (otherwise, the challenge would be irrelevant). The proponent can also choose to retract a strong commitment of her own. This is more difficult because she must also find and retract any other commitments that imply the retracted assertion (i.e., any reasons she may have given for her assertion). If there are inconsistent assertions in the proponent's commitment store, and the opponent challenges them, then the proponent must retract at least one of the conflicting commitments along with the reasons that led to it.

Walton and Krabbe (1995) also introduced an important distinction between *light-side* and *dark-side* commitments. Light-side commitments are based on explicit actions of the

participants of a dialogue, such as assertion or concession. The dark-side commitments of a participant, on the other hand, are the background beliefs of that participant. Dark-side commitments constrain some of the overt responses of the dialogue participants. Suppose the proponent, for example, challenges an assertion that the opponent believes is in the proponent's dark-side commitment store. The opponent may request that the proponent reconsider the challenge. If the assertion is in fact among the proponent's dark-side commitments, the proponent must withdraw the challenge. Rules such as this capture the role of critical thinking as a tool for improved understanding of one's own beliefs, and for bringing knowledge to bear on a problem that might otherwise have gone unused.

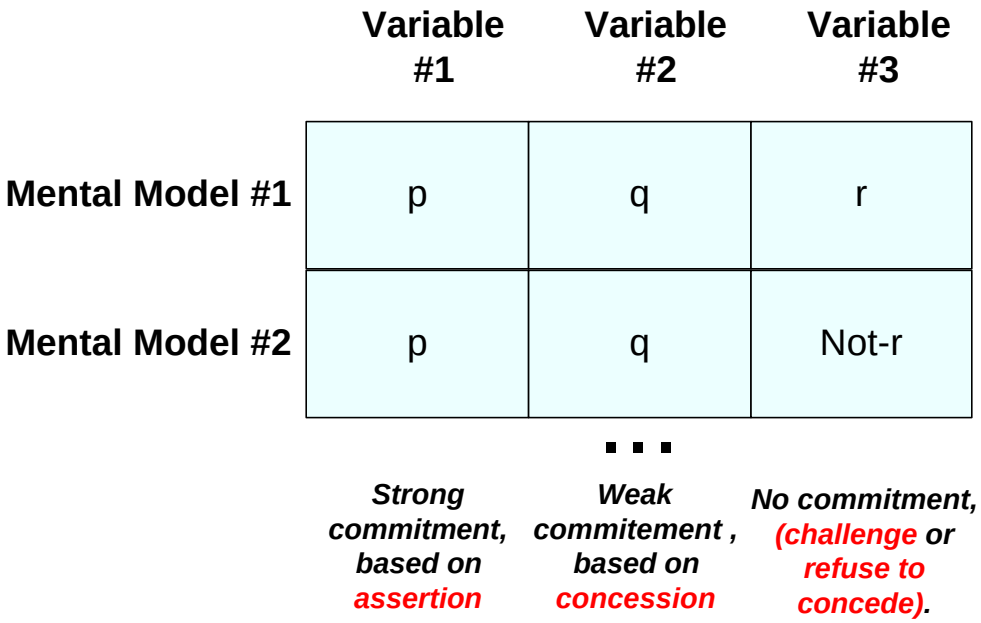


Figure 10. Three stances a participant in a dialogue might take toward a proposition, illustrated by propositions p, q, and r, respectively.

The light-side commitment store of a participant is her *explicit* situation model, i.e., a set of mental models corresponding to the participant's current understanding of the situation. Each mental model in that commitment store represents a state of affairs that is regarded as *possible* by a participant at that particular time. Figure 10 illustrates how a set of mental models is able to represent different kinds of commitments by participants in a dialogue. In this simple illustration, there are two possible states of affairs characterized by three propositional variables (p, q, and r).

p and q are each true in both mental models. Since there is no explicitly represented possibility in which they are false, they are *commitments* of this individual. The three dots under q, however, indicate that these are not the only possible states of affairs with respect to values of q. A fully fleshed out situation model *might* include states of affairs in which not-q was the case. Thus, q is only a weak commitment, e.g., a concession that is made to permit reasoning to proceed based on the supposition that q is true, until it is explicitly challenged. By contrast, the absence of dots under p indicates that all states of affairs where not-p is the case have been eliminated. p is therefore a strong commitment – for example, based on an assertion by the participant. Finally, r and not-r are each present in at least one of the mental models. r therefore is not a commitment at all. The representation of both truth and falsity means that the individual

has explicitly challenged, or expressed doubt about, r.[54] Figure 11 shows how different types of speech acts in a dialogue lead to changes in the commitment status of assertions.

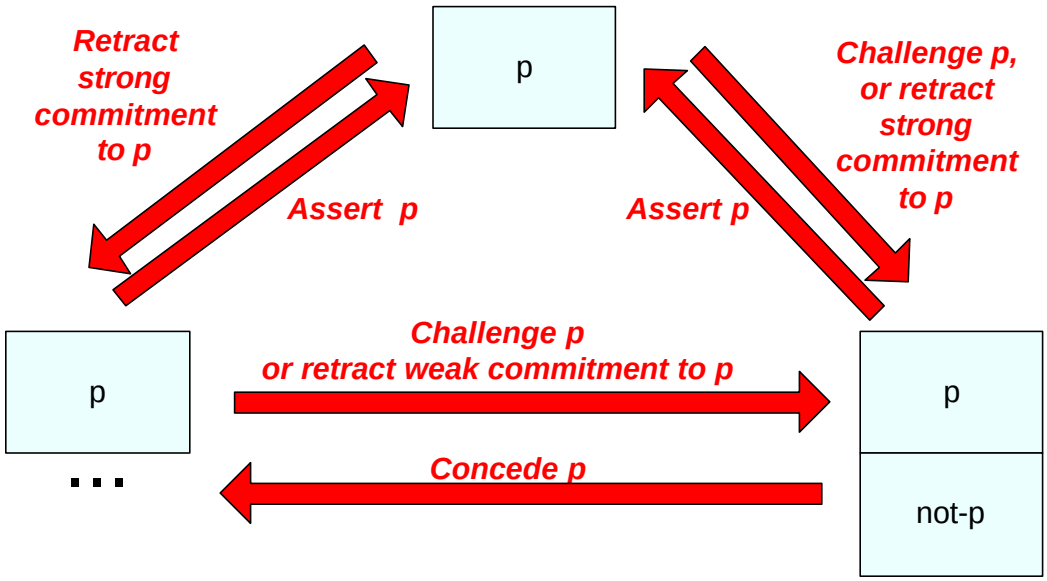


Figure 11. How speech acts in a simple dialogue change the commitment status of propositions for that speaker.

In a persuasion dialogue, the opponent tries to get the proponent to retract her initial thesis. The opponent's strategy therefore is to *add* plausible mental models to the proponent's situation model, and in particular, to add models in which the initial thesis is not true. The proponent, on the other hand, tries to get the opponent to concede the thesis. The proponent's strategy therefore is to reduce the number of plausible mental models in the opponent's situation model, and in particular to eliminate those possibilities in which her initial thesis is not true. Each uses knowledge, hunches, or inferences regarding the background belief system of the other to obtain useful concessions (Walton & Krabbe, 1995). The situation model of each participant will become more and more coherent with her background belief system as assertions are challenged and then reconsidered at the request of the opponent.

The Critic's Toolbox: Types of Challenge

A variety of different types of challenge and defense are available to the opponent and proponent, respectively, in a simple persuasion dialogue. In particular, an opponent may challenge the premises, inference, or conclusion of an argument. We will refer to these as Type A, Type B, and Type C challenges respectively, as shown in Table 6:

[54] The use of three dots in Figure 10 is an adaptation of Johnson-Laird and Byrne's (1991) device of placing three dots under the entire mental model. In Johnson-Laird and Byrne's use, the three dots show that the representation of an assertion containing logical connectives is not complete, i.e., that there are other, implicit mental models. We have extended this in the following respect: When the three dots appear under a specific variable, they indicate that other values of *that* variable may (or may not) occur in the additional, implicit mental models. The dots represent a weak commitment to the value(s) of the variable that are shown in the explicit models. By keeping the other models implicit, the participant *concedes* the explicit values for the purposes of argument.

Table 6. Different types of challenge in a dialogue.

	Type A	Type B	Type C
What is challenged?	Premises	Inference from premises to conclusion	Conclusion
Examples	Why do you believe your conclusion? What premises are you using to draw that conclusion?	Even if your premises are true, your conclusion might still be false [for the following reasons]!	Your conclusion is false [for the following reasons]!

In addition to the difference in "type," the challenge in each case may also be specific or general in nature. A specific challenge introduces a reason, i.e., a new variable that discriminates among mental models, while a general challenge asks the proponent to supply a reason. Challenges, therefore, are the engines that drive the elaboration of mental models and the application of more knowledge to the problem. A Type A challenge questions the evidential base for an assertion, either by asking for evidence (general) from the proponent or presenting evidence against the assertion to which the proponent must respond (specific). A Type B challenge questions an inference from evidence to conclusion, either by asking why the evidence is relevant to the conclusion (general) or presenting a reason why it is not relevant (specific). A Type C challenge questions the conclusion directly, either by asserting its negation (general) or by defending a particular way in which the conclusion is false (specific).

In sum, a *general* challenge demands that the proponent introduce a new variable that constitutes a reason: That is, the proponent is asked to discriminate one set of mental models (where the thesis holds) from another (where it might not), and show that the true mental model is in the set where the thesis holds. In a *specific* challenge, the opponent herself introduces the new variable to discriminate among mental models, and the opponent claims that the true possibility may be in the set where the thesis does not hold.

Table 7 characterizes each of the six types of challenge in more detail, by means of a paradigmatic interactive sequence. Each sequence starts with the precondition for that challenge, i.e., an assertion by the proponent that must have occurred previously in the dialogue. The challenge itself is shown underlined. Following the challenge are the types of responses available to the proponent. Note that for A and B challenges, the opponent does not make assertions but only points out possibilities. In response, the proponent must make assertions that rule out those possibilities. In Type C challenges, the opponent does make an assertion, and thus becomes the proponent of an opposing position.

Only one response is available to the proponent in the face of general Type A and Type B challenges. After a general Type A challenge, the proponent must respond with a reason for the conclusion. After a general Type B challenge, the proponent must respond with a reason the premise supports the conclusion, i.e., with additional information that clarifies the inferential connection between evidence and conclusion. This is a reason that works in conjunction with the original reason(s) to support the conclusion.

Table 7. Different types of challenge by an opponent (O) in a critical dialogue are underlined. Prior statements by the proponent (P) and possible responses by the proponent (P) are shown.[55]

	General Challenge (no specific reason)	Specific Challenge (cites a specific reason)
Type A **Challenge evidence for an assertion**	P: X is true. O: Why is X true?[56] P: Y supports X. Y is true.	P: X is true. O: Y supports not-X. Y might be true. P: Y is false. *Or* P: Z supports not-Y. Z is true. *Or* P: Why does Y support not-X? *Or* P: Y and Z do not support not-X. Z is true.
Type B **Challenge inferential relation between evidence and conclusion**	P: Y supports X. O: Why does Y support X?[57] P: Y and Z support X. Z is true.	P: Y supports X. O: Y and Z do not support X. Z may be true. P: Z is false. *Or* P: R supports not-Z. R is true. *Or* P: Y, Z, and R support X. R is true.
Type C **Challenge conclusion directly**	P: X is true. O: X is false. *P now acts as opponent to not-X.*	P: X is true. O: X is false. Y is true.[58] *P now acts as opponent to Y.*

[55] This taxonomy is a generalization of Rescher (1977, pp. 5-24), to which it adds Type C and general Type B challenges. X, Y, etc. are any statements, including negations.

[56] This implies that X might be false.

[57] This implies that even if Y is true, X might still be false.

[58] Y implies not-X, but not vice versa. Y is only one way in which X can be false. That is, Y is a contrary of X, but not a contradiction of X.

By contrast, Table 7 shows that the proponent has several options for responding to *specific* Type A and Type B challenges. In each case, the challenge itself has two parts: an assertion of a possibility plus an assertion about the implications of that possibility, if true, for a conclusion. The proponent can respond by challenging either of these parts. She can assert that the possibility is false and optionally provide a reason for its falsity. Or she can assert that the implication does not hold and optionally provide a reason why it does not hold.

An illustrative Critical Thinking Dialogue

To illustrate the challenges in a more concrete setting, we will take up a segment of the dialogue between MAJ Sud and MAJ Nord about location of attack (Figure 12). The example will show how the challenges further the aims of the dialogue and indicate how they are represented in the participants' evolving situation models.

MAJ Sud states a thesis (which happens to be a negative, i.e., *the enemy will not attack in the north*). MAJ Nord challenges rather than concedes MAJ Sud's thesis and thus takes the role of opponent. Note that in asking *Why[not]?*, the opponent is expressing doubt about a thesis and asking for a reason to support it, but is not *asserting* that the thesis is false. The demand for reasons for an assertion is a *Type A challenge*. When no specific reason to doubt the assertion is mentioned, it is a *general* Type A challenge.

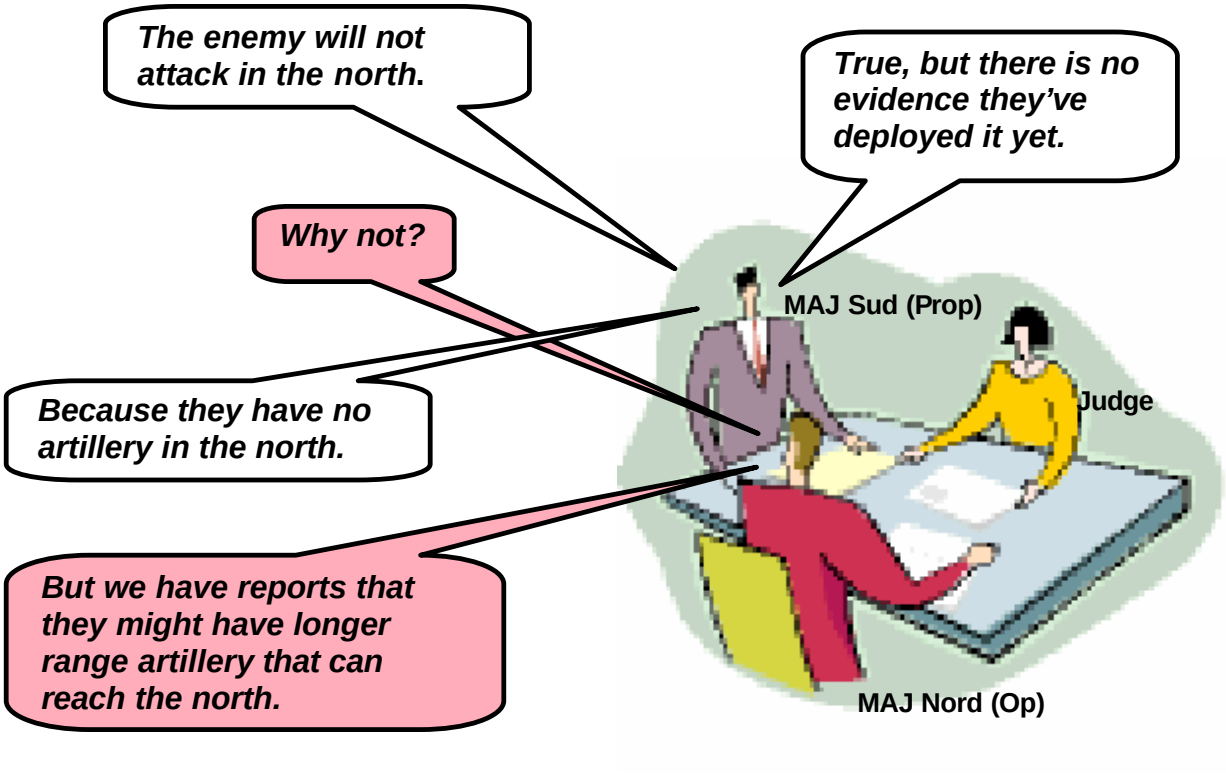
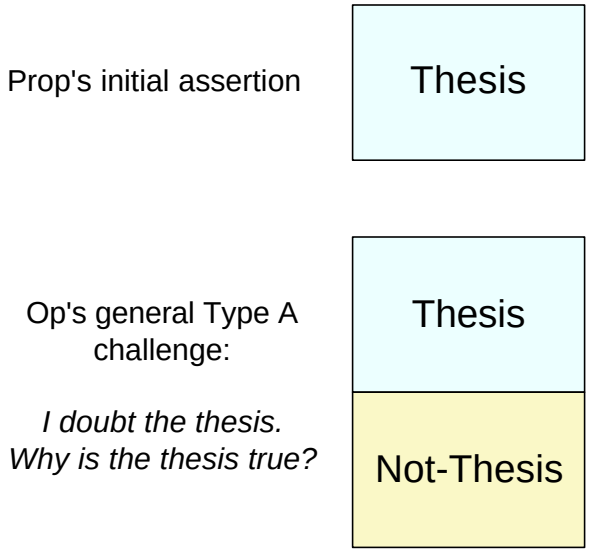


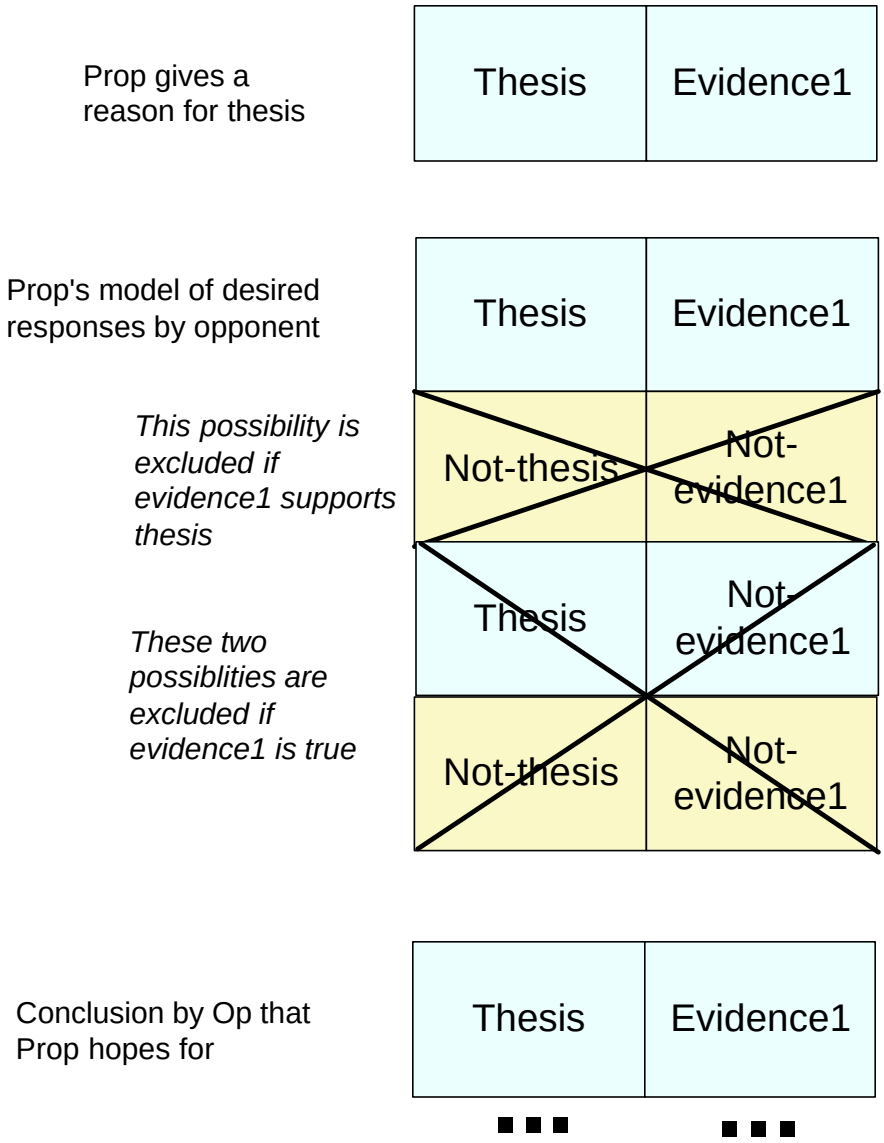
Figure 12. An illustrative critical dialogue, with MAJ Sud playing the role of proponent (*prop*) and MAJ Nord the role of opponent (*op*). Demanding a reason for an assertion (e.g., asking *Why?*) is a *general Type A challenge*. Challenging the inference from reason to conclusion with a defeater is a *specific Type B challenge*. Read clockwise from top left.

This brief exchange (up to *Why not?*) sets the ground rules of the subsequent discussion in two ways. First, (confrontation stage), a difference of views has surfaced. Second (opening stage), we assume that there is an implicit understanding that a persuasion dialogue will be conducted in order to resolve the difference in viewpoints. Moreover, at least for the time being, this will be a simple persuasion dialogue, in which MAJ Sud plays the proponent and MAJ Nord plays the opponent. The opponent will only express doubt, but will not (as yet) commit to a contrary thesis of his own. Thus, the explicit commitments of the two participants at the end of the opening phase of the discussion are as follows:



As the argumentation stage begins, MAJ Sud (the proponent) responds to MAJ Nord's Type A challenge (*Why not?*) by giving a reason: lack of artillery in the north. By introducing this new variable (location of artillery in the north), the proponent intends to narrow down the states of affairs considered possible by the opponent to one in which the thesis is true. In principle, of course, adding a new variable *adds* to the number of mental models, because there are more possible combinations of truth and falsity (four possibilities in the case of two variables). The end result, however, will be an elimination of models if (i) the opponent concedes that evidence1 supports the thesis (ruling out the combination that includes both evidence1 and not-thesis), and (ii) the opponent concedes that evidence1 is likely to be true (ruling out both possibilities in which not-evidence is the case). Thus, the opponent should concede the thesis. [59]

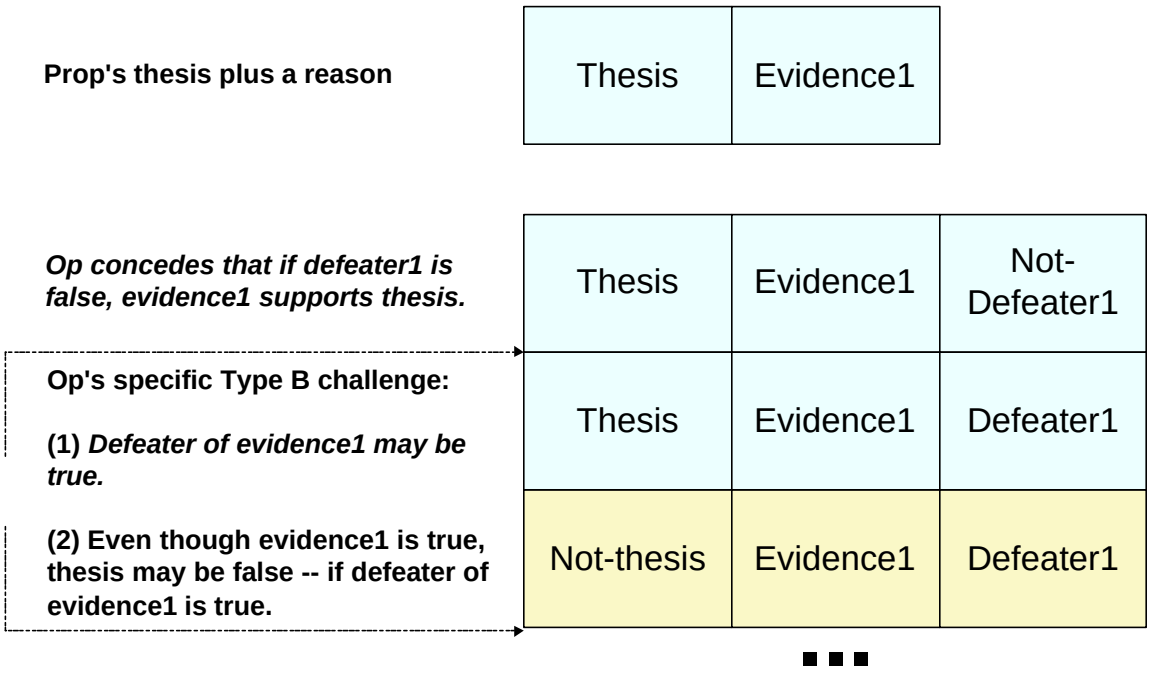
[59] The diagram shows that the logic of this reasoning requires four mental models. Experimental findings suggest that individuals will not in fact represent all four mental models (Johnson-Laird & Byrne, 1993). They represent a conditional, such as *If evidence then thesis*, by a single mental model: *evidence1 thesis*, unless it is necessary for the inference to flesh out other possibilities. In this case, it is not necessary, since the other premise (evidence1 is true) is also true in the model *evidence1 thesis*. Thus, they will go straight to the proponent's desired conclusion, that the thesis is true.



In his next move, the opponent concedes the truth of evidence1 (no artillery in the north), as the proponent expects. He also concedes that in general, evidence1 supports thesis. However, MAJ Nord challenges evidence1 as a reason for the conclusion (no attack in the north) in the presence of defeater1. The opponent brings up another new variable (range of artillery) and points out that there is a possible value of that variable (*long range artillery*) under which the combination of not-thesis and evidence1 is quite plausible.[60] Questioning the inference from evidence to conclusion is a *Type B challenge*. A *specific* Type B challenge includes a specific reason for doubting the inference, i.e., a *defeater* of the evidence. Even if evidence1 were true (i.e., no artillery in the north), the thesis might still be false (the enemy might attack in the north) – if the enemy had longer range artillery (defeater). The absence of artillery in the north in conjunction with the defeater (i.e., the development of longer range artillery) is no longer

[60] With longer range artillery, the enemy could locate it elsewhere to support an attack in the north.

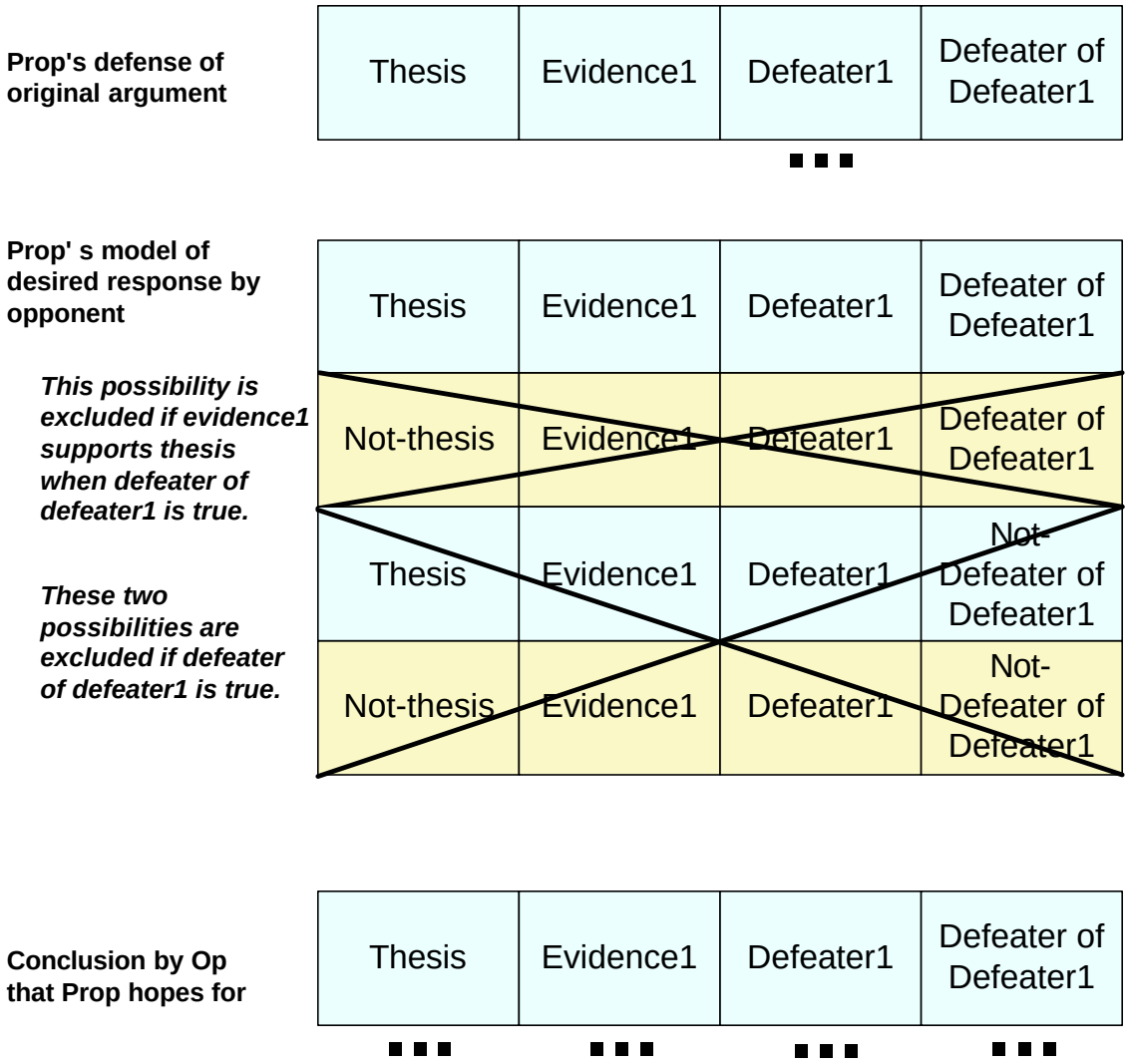
evidence for the thesis. The mental models of the proponent and opponent at this stage of the discussion are not what the proponent hoped for:



At this stage of the dialogue, MAJ Sud's only argument for his thesis (evidence1) has been defeated, and no conclusion about location of attack can be drawn. Unless he can come up with another defense, he will have to retract the thesis about no attack in the north.

The discussion thus far is a *simple* dialogue, in which only one of the parties (the proponent) is required or even allowed to make strong commitments. The burden of proof with regard to the thesis is on the proponent, while the opponent's role is merely to cast doubt on the proponent's argument. As a result, the opponent has not asserted anything on his own: First, the opponent merely concedes the truth of evidence1 (no artillery in the north), but is not committed to defending it. The three dots indicate that there may be other possibilities in which evidence1 is false, which he reserves the right to explore later. Second, the opponent does not commit himself to the defeater even though he is the one who brought it into the discussion. Simply raising the possibility that it is true is enough to create a doubt about the proponent's thesis. Third, a defeater may undermine an inference without providing evidence for the opposite conclusion. The opponent is not committed to the negation of the proponent's thesis, only to the possibility that it is false. In a persuasion dialogue, the proponent's goal is to get the opponent to concede the thesis, not strongly commit to it. Fourth, the opponent does not deny that the original evidence (absence of artillery in the north) supports the proponent's conclusion (no attack in the north) in general. Indeed, by introducing a defeater, the opponent implicitly concedes that the original evidence alone does support the conclusion. The defeater is a *distinction* between the situations where this support exists and the special cases where it does not (Rescher, 1977). In short, while absence of artillery may *generally* indicate no attack, there are circumstances (longer range artillery) in which this is not the case. Adding the possibility of longer range artillery thus neutralizes the support given by the proponent's evidence for the proponent's conclusion.

Defeasibility is an open-ended aspect of reasoning about the real world. Thus, the proponent, MAJ Sud, answers MAJ Nord's challenge by making a *finer distinction*. He concedes that the defeater may be true, but introduces yet another variable (reports on deployment of longer range artillery) in order to defeat the defeater. While the development of longer range artillery *generally* neutralizes absence of artillery as an indicator of location of attack, there are special circumstance where that is not the case. If *the enemy has not deployed the new artillery*, then mere development of the technology is irrelevant. With this additional discrimination, the original argument based on lack of artillery in the north regains its former force even if the defeater (long-range artillery) is true. Here is the proponent's model of how he hopes the opponent will respond:



The three claims in combination (lack of artillery in the north, and possession of longer range artillery that has not been deployed) provide support for no attack in the north. When evidence1 is true and the defeater of the defeater is true, the thesis is likely to be true – whether or not the defeater is true. Thus, the proponent once again aims to convince the opponent to eliminate his second mental model. Since the thesis is true in the surviving possibility, MAJ Sud's original

conclusion has been vindicated – unless of course the opponent comes up with another challenge.

Up to now, MAJ Nord has been content to express doubts about MAJ Sud's position and arguments. MAJ Nord's situation model has included mental models in which both the thesis and not-thesis were the case. However, MAJ Nord can challenge MAJ Sud's claim more strongly, by asserting a contrary thesis of his own (Figure 13). If he does, there is an implicit shift in dialogue type from simple to *compound*. The roles of proponent and opponent become symmetrical. Each participant must defend his own thesis and attack the thesis of the other party. Attacking the conclusion directly by the assertion of a contrary thesis is a *Type C challenge*. When the contrary thesis is the negation of the original thesis (i.e., not-thesis) as in this case, it is a *general* Type C challenge, because it does not specify in any detail *how* the thesis is false. If a more specific thesis is asserted (e.g., *attack will be in the south*), which is incompatible with the original claim but is not merely its contradiction, we have a *specific* Type C challenge. As we shall see, the shift from simple to compound persuasion dialogue (via a Type C challenge) introduces some significant wrinkles in determining which side has "won."

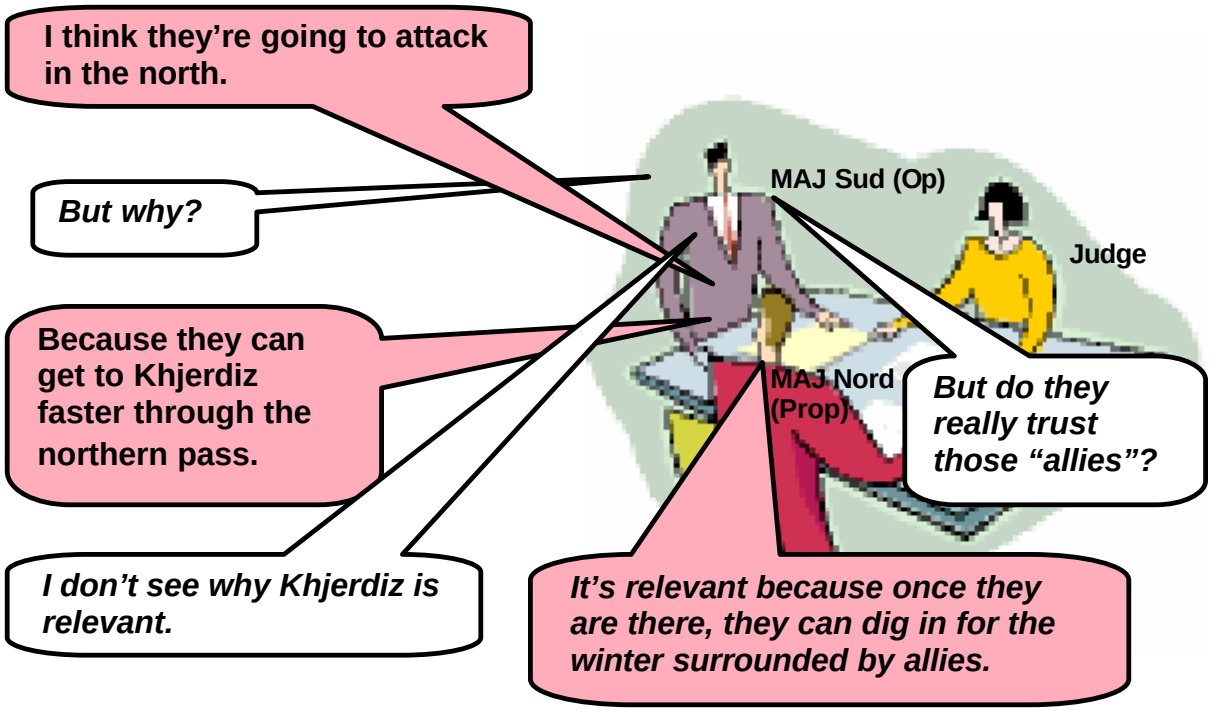
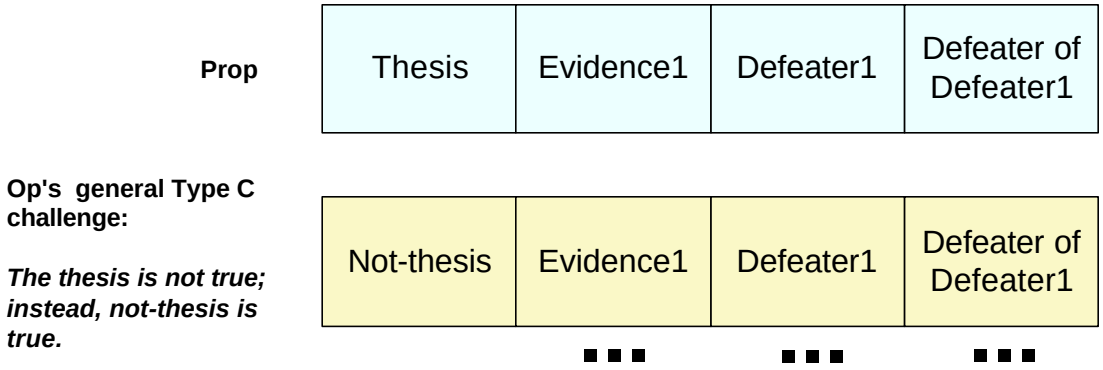


Figure 13. The former opponent (MAJ Nord) asserts the contradictory thesis, a general Type 3 challenge. MAJ Sud issues a general Type B challenge by questioning the relevance of the evidence. Read clockwise from top left.

After MAJ Nord's assertion that the attack will be in the north, both participants have a single mental model:

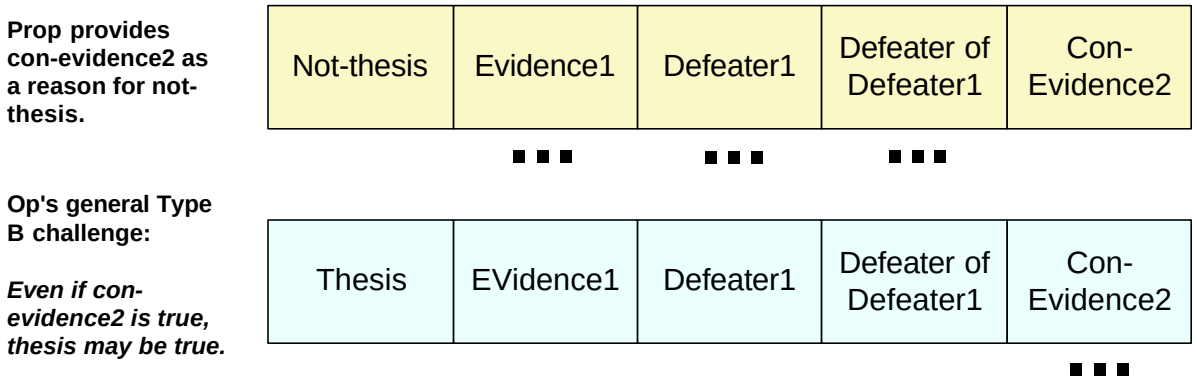


MAJ Nord's position as it stands is unstable. The problem is that there are no *other* explicit differences between the two participants in this dialogue. MAJ Nord has conceded everything that MAJ Sud has said, including its relevance to the thesis – except the thesis itself. He must concede the thesis as well unless he can produce a further discrimination that explains why thesis is not supported. He must probe his background beliefs to come up with something else: either another challenge to MAJ Sud's position or else a reason for not-thesis. MAJ Sud in fact asks him to produce such a reason with a general Type A challenge (*But why?*).

MAJ Nord responds that the enemy can get to a certain town faster through the northern pass. (We call this con-evidence2 to indicate that it supports the contrary of the thesis, viz., not-thesis.) MAJ Sud's first response to this assertion is to ask why it is relevant. A question about the relevance of purported evidence is a *general Type B challenge*. It probes the reasoning from premise to conclusion, rather than the premise or the conclusion directly. The challenge says, in effect, *I believe it is possible for both con-evidence2 and thesis to be true. So, why does it follow that if con-evidence2 then not-thesis?* By contrast, a *specific* Type B challenge, as we saw earlier, introduces an explicit defeating condition, i.e., a specific reason to doubt the inference.

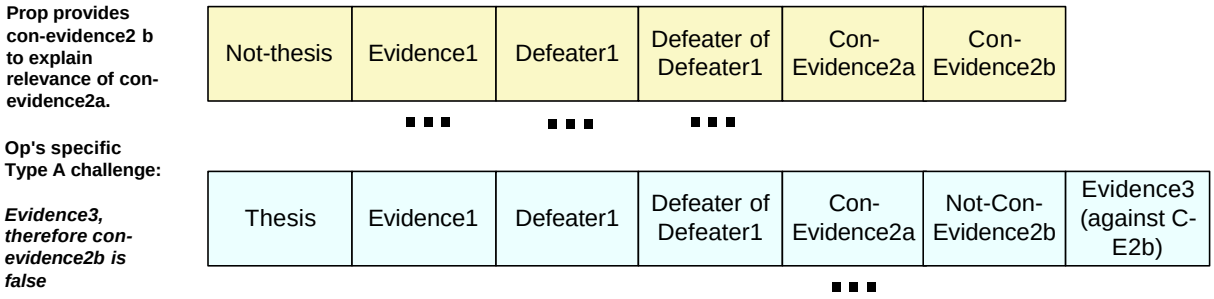
Since MAJ Sud (now acting as the opponent) does not accept the reasoning by means of which con-evidence2 is supposed to eliminate the thesis, his mental model contains both the thesis and con-evidence2:[61]

[61] In this segment, MAJ Nord plays the proponent and MAJ Sud plays the opponent. We have reversed the order of their respective models since MAJ Sud is now responding to MAJ Nord rather than vice versa.



MAJ Nord responds with an additional reason, that the enemy will be safe in the town over the winter. This reason is not intended to support not-thesis independently. Rather, it explains the relevance of con-evidence2. It works *in conjunction* with con-evidence2 to support the conclusion. An attack in the north is supported by the combined information that it is easier to reach the town through the north and that the enemy will be safe in the town. Because of this linkage, we will re-label the first reason (faster to get to the town) as con-evidence2a and the second reason (safety of the town during the winter) as con-evidence2b. The two in combination make up a single line of defense for not-thesis. Since neither con-evidence2a nor con-evidence2b is very effective without the other, MAJ Sud only challenges one of the two in order to defeat or at least weaken MAJ Nord's argument. MAJ SUD offers a specific Type A challenge against con-evidence2b – a reason to think that the enemy might not want to spend the winter in that particular town after all (*they may not trust their alleged allies*).

We will stop here. The explicit situation models are now the following:



Handling Defeasibility

Informal logicians, psychologists, philosophers, and artificial intelligence researchers generally agree that most real-world inferential conclusions are *defeasible,* i.e., subject to defeat by new information. Nevertheless, defeasibility is not handled well within either formal deductive logic or informal logic. Formal logicians deal with defeasibility by tinkering with the premises and informal logicians deal with it by tinkering with the inference rules of a reasoning system.

Formal logicians handle defeasibility by adding the falsity of the defeater to the antecedent of a conditional that serves as a premise in the argument – e.g., *If there is no artillery in the north and longer range artillery has not been developed and deployed, then there will be no attack in the north.* A problem with this tactic is that it blocks reasoning with incomplete information. After the underlined clauses are added to the conditional premise, artillery location alone will not be available as an indicator of location of attack. But in many circumstances, it is neither possible nor worth the time to find evidence on all possible defeaters. As the conversation between MAJ Sud and MAJ Nord continues, more exceptions and exceptions to exceptions may be brought forward. Each new discrimination in the antecedent of the conditional ratchets up the demand for information before the inference can be regarded as valid. As a result, the decision maker might never be able to reach a conclusion at all.

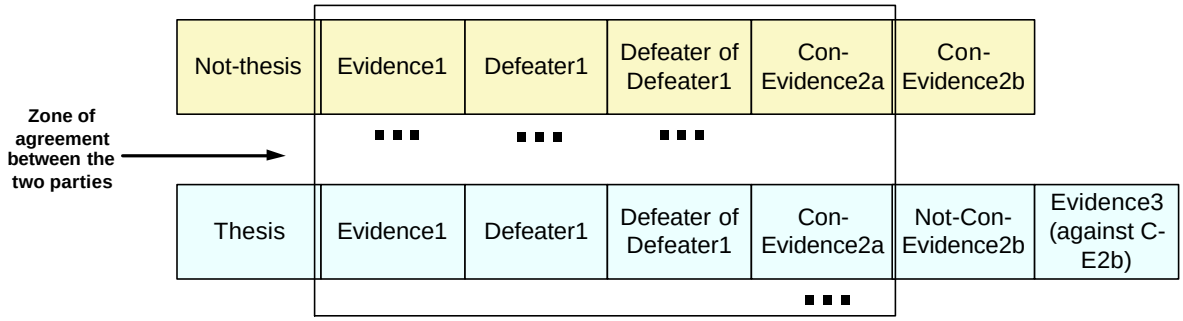
A solution more in line with informal logic is to develop an alternative reasoning framework by adding special default inference rules. According to these rules, a conclusion follows as long as there is no positive evidence that the defeaters are true (e.g., Reiter, 1980). They are presumptively false, hence, it is not necessary to *show* that they are false. This approach permits inference under conditions of incomplete information, but it has the opposite problem. It rules out the possibility that on some occasions we might in fact want to require positive evidence that a defeater was not the case before rushing into a conclusion – e.g., when the costs of errors are high and time is available to collect information and think. Another problem for both approaches is that the list of potential defeaters is indefinitely long, and advance specification of all defeaters in default rules or conditional deductive premises may be impossible even in principle. The set of defeaters for the inference from an effect to a cause, for example, must include *all* the other possible causes. Finally, neither the formal nor the informal approach provides guidance or flexibility in determining how long the process of generating defeaters and collecting information about them should go on. Proficient decision makers are able to adapt the reasoning process to specific circumstances, to act decisively on a subset of the relevant information in situations where that is necessary, and to demand more thinking and more information where that is called for.

The problem of defeasibility invites a constructive solution involving a synthesis of mental model theory, dialogue theory, and reliability. Defeasibility always involves an incomplete set of mental models. A defeater (or a defeater of a defeater, and so on) reflects the introduction of a new variable to the discussion, and this in turn stimulates consideration of possible states of affairs that were overlooked but which are relevant in the current context. Defeasibility therefore lends itself to a mental model-based approach that directly represents the alternative possibilities and the variables by which they are discriminated (Johnson-Laird & Byrne, 1991; Johnson-Laird, Legrenzi, Girotto, Legrenzi, & Caverni, 1999). Dialogue theory provides norms for the process of challenge and response during which mental models are elaborated and accepted or rejected. Finally, judgments of reliability determine what process should be used and when the process should stop as a function of external variables like stakes and available time.

Expanding Knowledge and Sharing Mental Models

As the critical dialogue progresses, new variables are added to the explicit situation model, either to challenge or to defend the proponent's thesis. Each new discrimination brings more information into the conversation, and has the potential both to deepen understanding and improve the predictive accuracy of the participants' situation models. Moreover, every

discrimination that is proposed by one party and conceded by the other increases the degree of overlap among their situation models, hence, the amount of shared situation understanding. Thus, while the explicit goal of critical dialogue is to arrive at the most plausible resolution of a difference of opinion, its secondary accomplishments can be just as significant. For an individual, critical dialogue elicits knowledge that may not otherwise have been used in the current problem, and may lead to creative insights that have application in other situations. Among members of a team, critical dialogue expands the sharing of knowledge about the problem and about the domain. In this example, the participants have a fairly substantial zone of agreement on the facts about the current situation. Perhaps more importantly, they have acquired a shared understanding of what the relevant issues are.



Each variable is a dimension along which possible states of affairs can vary. Thus, each new variable increases the number of *logically* possible situations, i.e., the combinations of truth and falsity of the propositions corresponding to the variables. For example, since there are six variables (columns) in the final step of our example, there are actually $2^6 = 64$ possible states of affairs. It would be most difficult for humans to keep that many mental models in mind, and fortunately it is not necessary. As the example illustrates, because of the role of background beliefs and the avoidance of explicit deductive inference, the actual number of mental modes that needs to be considered is much lower, and does not necessarily increase much at all as new dimensions are introduced. This example required explicit representation of at most three of the 64 logically possible mental models at any given time. The objective of the proponent, after all, is to *reduce* the number of mental models until all the survivors contain the conclusion, and she does so by introducing new considerations that interact appropriately with background knowledge to eliminate possibilities. In addition, concessions by the opponent function as assumptions which reduce the range of alternatives that need to be considered. Finally, according to Walton (1998), assertions that are appropriately based on argumentation schemes have a presumption of truth and can be challenged only when here is a specific reason to doubt them.

Argumentation Schemes and Causal Structure

Con-evidence2a and con-evidence2b in combination appear to be an example of what has been called a *linked* argument by informal logicians. There are numerous attempts to define what is meant by a linked argument, none particularly successful (Walton, 1996b). One problem is that even when evidence items seem to be linked, falsifying one can leave the other with some evidential value. In this example, being a fast way to get to the town (con-evidence2a) might be relevant as evidence for attack in the north even if con-evidence2b is false, as long as there are other possible reasons for wanting to be in the town than stated by con-evidence2b. The real

linkage seems to be at a more abstract level, e.g., between (i) Y is a means to achieve X, and (ii) X is a desirable outcome – where Y is attack through the north and X is being in the town.

This more abstract level of representation supports Walton's (1996b) proposal that linked arguments must be defined in terms of argumentation schemes. The role of argumentation schemes need not be entirely conventional, but may be largely determined by causal structure. In particular, the argumentation scheme provides a *Markov blanket* around the variable of interest, i.e., the conclusion, attack through the north. That is, the variables introduced by the argumentation scheme "block" the influence of all other variables on the conclusion. If we know that (i) and (ii) are the case, we do not need to further consider in what ways Y might be a means to achieve X, or in what ways X is a desirable outcome. Thus, the argumentation scheme provides the critical questions that must be asked in order to challenge the conclusion, and these are directly "linked" to one another. These questions may then lead to other questions, which are only indirectly linked.

Winning and Loosing

If the discussion must come to an end (e.g., because action is necessary), who has "won"? Should the thesis (no attack in the north) be accepted or rejected? Notice that throughout the illustrative exchange, in order to defend his thesis against challenges by the opponent, the proponent had to make commitments – first, to evidence1, then to the defeater of the defeater of evidence1. The opponent can "win" if her challenges force the proponent into commitments that are implausible and difficult to defend (Rescher, 1977, p. 23). Thus, the opponent's strategy is to find weak points to challenge in the proponent's situation model. She increases the number of mental models to force the proponent to consider and respond to alternative possibilities in which her conclusion is not the case. If the proponent cannot counter a challenge (e.g., by a new discrimination), the opponent wins.

Conversely, if the opponent cannot find any claim to challenge, then the proponent wins. The proponent's strategy therefore is to steer toward firm ground, to end a chain of reasons or defeaters of defeaters with assertions that are presumptively true, or at least highly plausible, in the current context (Rescher, p. 44). *Presumptive truths* are assertions arrived at in accordance with an *argumentation scheme*, i.e., a framework that distinguishes premises, a presumptively true default conclusion, and possible critical questions (Walton, 1996a). Argumentation schemes, according to Walton, are normatively binding kinds of reasoning, i.e., appropriate moves and countermoves in a dialogue. Each argumentation scheme is associated with (i) typical premises and a presumptive (default) conclusion, and (ii) a set of critical questions that may be used by the opponent to challenge the conclusion. The key feature of argumentation schemes is that they reverse the usual burden of proof. A proponent who uses an argument scheme correctly has presumption on her side.

In the absence of either "I give up" outcome, the success of the proponent must be judged in terms of the *plausibility* of the assertions to which she is committed but which the opponent has not conceded. In this example, there is only one, the defeater of the defeater. It also makes sense, in judging the outcome, to take into account that some of the opponent's "concessions" may be due to lack of opportunity to challenge (Rescher, p 23), and some of the proponent's unjustified commitments may be due to lack of time to defend. (For example, given more time, the opponent might eventually have chosen to challenge rather than concede evidence1.)

The Role of Reliability

A problem that is not addressed by either mental model theory or dialogue theory is the choice of a strategy that will reliably achieve *external* objectives. This gap exists because of the *internalist* character of mental model theory and the quasi-internalist character of dialogue theory. According to internalist theories, criteria for assessing the acceptability of beliefs must always refer to cognitively accessible internal representations, and not external facts of which the cognizer was not aware. Dialogue theory refers to two people engaged in an overt verbal exchange. Despite this public character, dialogue theory has significant kinship to internal approaches. It focuses primarily on internal conformity of a verbal exchange to the norms of a particular type of dialogue, rather than on the selection of the dialogue type and regulation of the dialogue itself in a way that is appropriate for an external task. Moreover, the norms themselves are referred to as conventional frameworks that derive justification from shared expectations. Two features clinch its internal status: First, the norms are applied only to facts that are *known* to one or both of the participants.[62] Second, the evaluation focuses on proximal or internal objectives associated with a particular type of dialogue, e.g., resolving a conflict of opinions, rather than on distal or external objectives, such as accomplishment of a task or mission. Because of these internal norms and proximal objectives, dialogue theory tends to describe self-enclosed games. Its internal focus is responsible for the failure of dialogue theory to adequately address three key issues: The selection of the appropriate types of dialogue, the rules for bringing a dialogue to an end, and how to determine the winner of a persuasion dialogue. All of these issues require judgments of external reliability.

Dialogue theory does not (thus far) address the reasons for choosing a particular dialogue type on a particular occasion, i.e., how different types of dialogues, such as negotiation, inquiry, persuasion, information seeking, deliberation, and quarrel, might be conducive to the accomplishment of different real-world objectives (Walton, 1998). The same dialogue type and sequence of moves might be judged appropriate in one context but not in another. An expert-consultation dialogue might make sense when one participant has significantly more knowledge and experience than the other; but an information seeking dialogue should be used when one party merely has information that the other party lacks. Should interactions between team leaders and team members be based on negotiation (members are rewarded for desired behavior) or persuasion (the team has a shared set of beliefs and values)? Should interactions between team leaders and team members be one-sided (i.e., simple persuasion dialogues in which only one party asserts a position) or two-sided (team members are permitted to assert and defend their own positions)?

Dialogue theory does not provide an adequate solution for when to stop a dialogue. For example, in the critical discussion that we looked at above, there was no limit to the number of challenges and responses, hence, to the number of features and alternative mental models that might be considered. Participants need to know when challenges should come to an end and the current best conclusion acted upon, and this usually depends on external context. For example, the same dialogue might justify acceptance of a conclusion when there was limited time or

[62] Rules pertaining to dark-side commitments are an exception. A participant does not know with certainty what her dark-side commitments are. Thus, she cannot be absolutely sure that she is following the rule that says she must retract challenges that conflict with those commitments. She does not have access to all the information that is relevant to fulfilling this requirement.

information to make a decision, but might be insufficient to justify a conclusion when more information or more time is available. The costs of incorrect conclusions might also influence the amount of time devoted to the dialogue.

Dialogue theorists address the issue of winning and loosing in terms of clear-cut cases, in which either the proponent retracts her original assertion or the opponent withdraws her challenge. Real cases may not always be so easy. Time constraints may bring a dialogue to an end before definitive closure is achieved. In such cases, it is necessary to determine which position was superior at the time the dialogue came to an end, taking into account the opportunities that the participants had to challenge one another. A somewhat deeper problem is that there may be considerations favoring each side, and the final determination will require a delicate evaluation of the remaining undefended assumptions on each side. This requires judgments about the relative reliability of different belief formation processes as well as the coherence of the alternative mental models both internally and with respect to background beliefs.

According to van Eemeren & Grootendorst (1992), decisions of these kinds take place during the opening stage and the concluding stage of the dialogue, rather than during the argumentation stage. For example, the type of dialogue should be agreed upon between the participants at the beginning of the dialogue, and the concluding stage determines when the dialogue ends and who won. Segregating them into different stages suggests that these decisions are qualitatively different from argumentation proper. But dialogue theorists do not address how the decisions should be made. Placing them in different temporal stages is quite artificial and only makes matters worse, since it eliminates the possibility of continuous review of the dialogue based on new information acquired during argumentation. Such information might lead to a shift from one type of dialogue to another (Walton, 1998), or it might change the estimation of how the risks of further delay balance out the costs of an incorrect conclusion, and thus affect the decision of when to stop. A more promising direction is to introduce an *externalist* point of view, which can exist more or less in parallel with the internalist perspectives of the opponent and proponent. The externalist point of view takes into account likely outcomes and their associated impact on objectives.[63]

To help dialogue theory bridge the gap between internal and external concerns, it is convenient to provide a third role, that of a *judge,* in addition to those of proponent and opponent (Figure 14).[64] All three of the issues just discussed belong among the duties of the judge. The judge evaluates the reliability of alternative types of dialogues for the current context and purposes. The judge evaluates the status of the argument at any given time to determine the most plausible current position, i.e., the winner if the dialogue were to end at that moment. And finally, the judge continuously weighs the value of continuing a particular dialogue versus the value of stopping and committing to the most plausible current position.

---

[63] Note that if the proponent or opponent in a persuasion dialogue appealed to outcomes, it would be considered a fallacy! One should not introduce any element of negotiation or bargaining in deciding issues of truth. But such considerations *are* relevant in determining the type and duration of the dialogue itself.

[64] van den Hoven (1987) also introduces the role of judge to account for external justification.

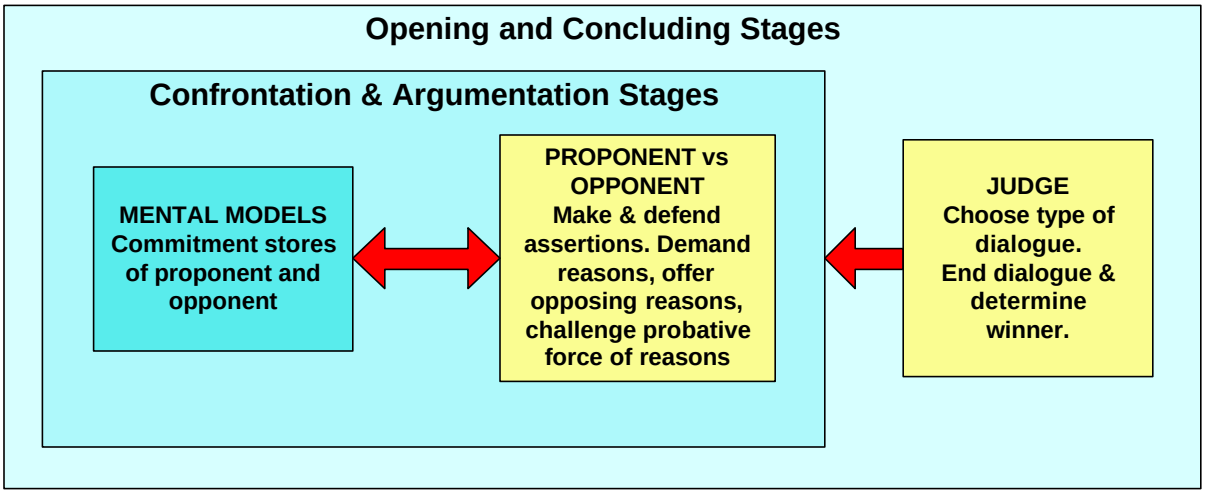


Figure 14. Three part model of critical thinking in terms of stages and roles in a critical dialogue.

Figure 14 shows that each component of the critical thinking model (Figure 9) corresponds to a dialogue theory concept. As we have seen, mental models correspond to the commitment stores of proponent and opponent. Critical dialogue corresponds to the argumentation between proponent and opponent in which the mental models are evaluated and improved. The judge determines the reliability of different processes and regulates them accordingly. This includes selecting the belief formation process that is most reliable in the current context (e.g., rapid recognition versus recognition plus critical thinking), and determining when the output of the process is sufficiently reliable to terminate it. In performing these functions, the judge is subject to the same capacity limitations as the proponent and opponent (especially if all the roles are played by the same cognizer). As a result, the Judge will not generally *optimize* strategy choices. Rather, in accordance with the principles of bounded rationality (Simon, 1997; Gigerenzer & Selten, 2001), the judge will become adapted through experiences of success and failure in the use of various cognitive processes and mechanisms in different contexts. The judge may use relatively automatic processes to select and regulate belief forming strategies, or may evaluate the reliability of different strategies by explicit reasoning. The common core of the judge's functionality is judgment about the trustworthiness of a cognitive faculty from a standpoint that is external to that particular faculty.[65]

Figure 15 shows how justified conclusions depend on the reliability of dialogue-related processes and traits over different time spans, as shown more generally in Figure 6. The overall reliability of the outcome may depend on the reliable functioning of general critical thinking traits over long periods of time, on the selection of reliable strategies for conducting the specific roles in the dialogue, and on the reliable execution of those strategies in accordance with the appropriate dialogue rules.

[65] The external perspective might use the very faculty that is, in other respects, under evaluation. It is an unavoidable fact of the human condition that we must use reasoning to evaluate reasoning. The difference in perspective consists in the fact that we can use reasoning *about* dialogue types to assess the expected reliability of the reasoning that takes place *within* a particular dialogue.

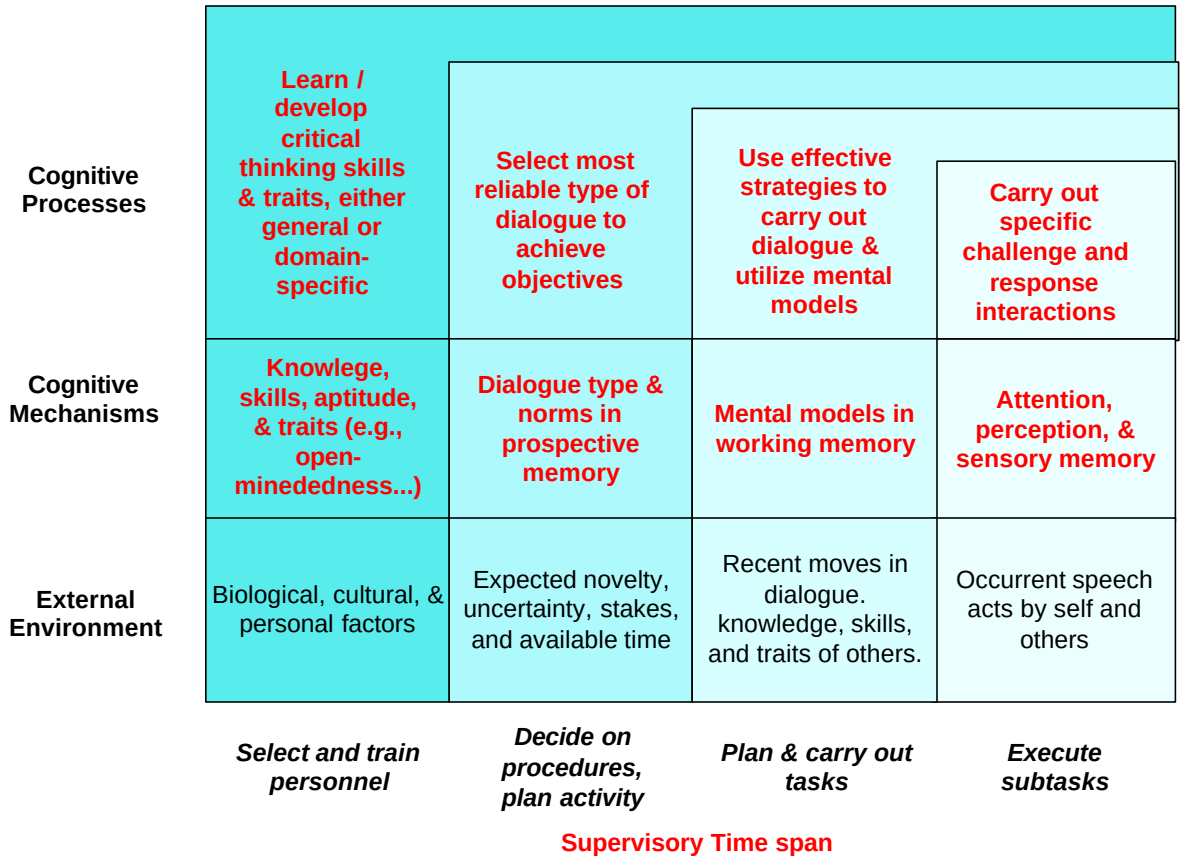


Figure 15. A cognitive model of critical thinking mechanisms, processes, and environmental factors operating over different time spans.

Critical thinking in the strongest sense involves all three levels, as shown in Figure 14. But the introduction of a reliability-based judge generalizes critical thinking beyond the evaluation of explicit reasoning or critical dialogue. Other belief-generating faculties, such as perception, recall, and recognition can also be assessed critically in terms of their reliability, even though they do not themselves involve reason-giving and critiquing. Thus, there is a weaker but still very important sense of critical thinking in which the judge evaluates not only the reliability of different dialogue types, but more generally, the effectiveness and efficiency of alternative cognitive faculties and processes. In some situations, taking time to reason may not be the best solution.

8. FROM ARGUMENT TO DIALOGUE AND STORY

In this and in the following chapters we will make a case for the critical thinking model. The case is based to a great extent on limitations of the traditional view of reasoning as exclusively or primarily *argument*. The evolution of critical thinking that we envisage moves from foundationalism (an argument-centered approach reflected in mainstream formal and informal logic) to (i) coherentism (a mental model, story, and dialogue-based approach) and (ii) reliabilism (as reflected in real-world oriented naturalistic and adaptive decision making). At the same time we will clarify further how the components of our theory work together in critical thinking and what the actual role of argument is.

Dialogue and the Architecture of Belief

Siegel (1997) falls within the internalist tradition when he says that

> …being a critical thinker requires basing one's beliefs and actions on reasons… the beliefs and actions of the critical thinker, at least ideally, are *justified* by reasons for them which she has properly evaluated (p.14; italics in original).

This view appears everywhere in the critical thinking literature, to the point where it may seem to be little more than simple common sense. Justification of beliefs by explicit argument is a central demand in virtually all textbooks and theoretical discussions of critical thinking.

If reasons are themselves beliefs, then Siegel's (1997: p. 16) principle is problematic even as an ideal. The demand that every belief be justified by an argument based on other beliefs leads to an infinite regress of *Why?* questions (Dancy & Sosa, 1992: p. 209-212), or Type A challenges. Reasons must be provided to justify the reasons, additional reasons to justify the reasons of those reasons, and so on. Critical thinking may never come to an end. There are only four ways to avoid such a regress within the internalist tradition: The list of reasons is infinite, it circles back on itself, or it stops. If the list stops, the reasons at which it stops may be justified or unjustified. If justified, justification must be due to intrinsic properties of the beliefs, not inferential relations to other beliefs.

If the list of reasons continues down infinitely without ever reaching bottom, conclusions can never be justified. The result is *skepticism* about the possibility of knowing anything (e.g., Unger, 2000; Foley, 2000). A closely related alternative is to end the list of reasons at an arbitrary, unjustified stopping point. This is the *relativist* position, that beliefs are not justified absolutely, but only relative to assumptions that happen to be accepted in a particular domain or culture, or by a specific individual at a specific time and place. Some critical thinking theorists come close to endorsing this view. For example, McPeck (1994; p. 109) states that "not only are canons of validity different, but what might be fallacious reasoning in one context or domain, might be perfectly correct in another." On this view, there may be no shared criteria of belief acceptance across different communities, domains, or cultural contexts. If there are, the shared components may be insufficient to support rational discussion – not good news for critical thinking theory!

A third possibility is that the chain of reasons eventually repeats. For example, continuing down the chain of reasons, we would eventually arrive again at the conclusion with which we began. Siegel, like many critical thinking theorists and informal logicians, rejects the idea that a

chain of arguments can legitimately circle back on itself (e.g., Siegel, 1997: p. 71). Such an argument commits the fallacy of *begging the question*, in which the reasons for a conclusion turn out to contain the conclusion itself. In other words, the reason for accepting *p* is, ultimately, *p* itself. The justification of a belief depends on the justification of that same belief.

Siegel, along with most other theorists in critical thinking and informal logic, is therefore committed to the fourth possibility, that the chain of reasons must come to rest on solid ground, with evidence that is not inferred from other beliefs, and which can serve as a *foundation* for inference of other beliefs. This view is called *foundationalism* (Chisholm, 1977; Pollock & Cruz, 1999), and it rejects an assumption common to the other three responses. Skepticism, relativism, and coherentism all view justification exclusively in terms of inferential *relationships* among beliefs. They assume that a belief cannot be justified on the basis of its own intrinsic properties. For foundationalists, this will not do. Inference cannot generate justification out of thin air; it must *transmit* justification from beliefs that are already justified in some other way. Inference must eventually be grounded in intrinsically justified premises.

Foundationalism has evolved from a classical version to a contemporary version, and the latter underlies most of the current work in informal logic and critical thinking. Another view, called *coherentism*, emerged in reaction to failures of both versions of foundationalism to deal with uncertainty, and is a more sophisticated, non-skeptical version of the circular reasoning option (in response to the threat of infinite regress). We shall see that these three views can each be understood along two parallel planes – in terms of different restrictions they place on the opponent or critic in a dialogue, and in terms of the type of belief architecture that results from, and accounts for, the relevant type of dialogue.

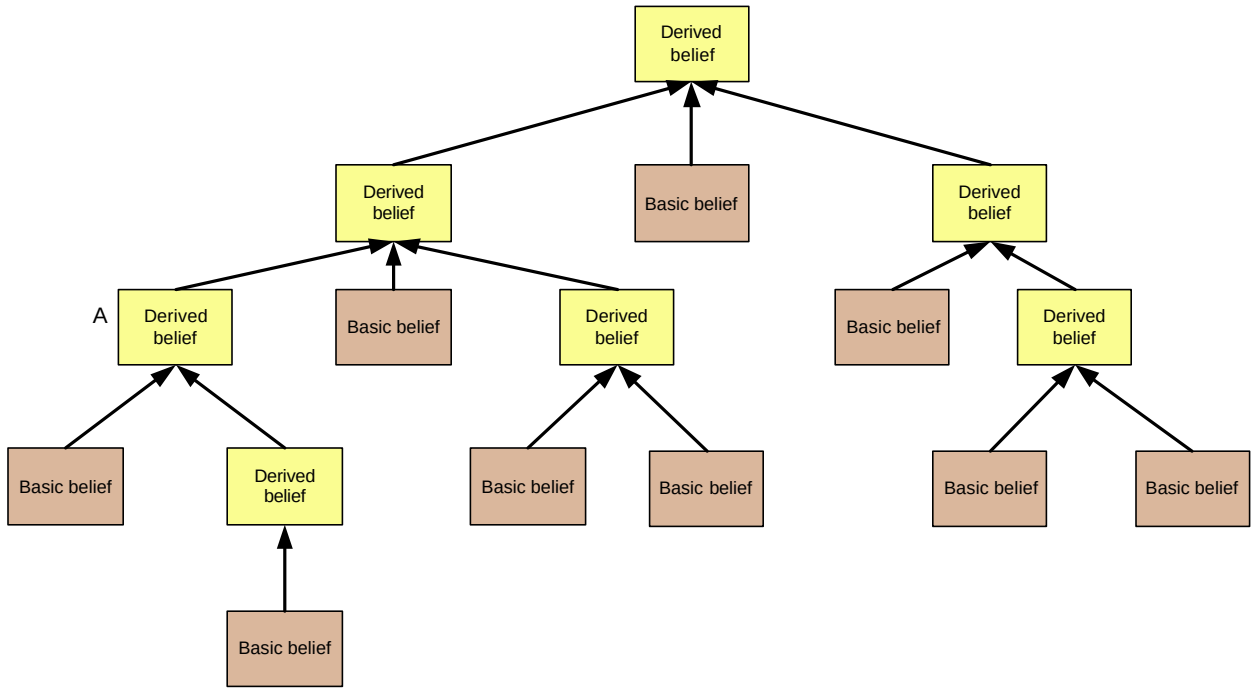


Figure 16. Foundationalist paradigm for acceptability of beliefs: a pyramid. Arrows represent arguments. Every chain of argument must be traceable back to basic beliefs (shaded boxes) at the bottom of the pyramid.

Is Critical Thinking Rigorously Based on Solid Foundations?

Foundationalism is an internalist theory that makes two demands on cognizers: (i) to discriminate two different kinds of beliefs, basic and derived, and (ii) to utilize a single direction of inference, from the first (basic) to the second (derived). The resulting belief system is a pyramid, as shown in Figure 16, in which every belief is logically derived by argument from beliefs at the level below, except basic beliefs at the bottom which are known directly (Sosa, 1991: pp. 19-34). Basic beliefs are where the chain of *Why <conclusion>?* questions must come to an end (see Type A above). *Classical foundationalism* asserted that such a pyramid could confer *certainty* on all its components.[66] Basic beliefs are known with certainty due to intrinsic properties, e.g., because they report the immediate evidence of the senses (*I seem to see a tank*), introspection (*I feel pain*), or logical truths (*that is either a tank or not a tank*). If a belief is not intrinsically justified in this way, it must be justified relationally, by argument based on other beliefs, i.e., by answering the *why* question. In each argument, moreover, truth of the grounds must guarantee the truth of the conclusion, by deductive logic. The classical foundationalist view thus leads to the following normative definition of critical thinking (a special case of the internalist definition given previously):

---

[66] From the point of view of internalism, there is some inevitability in the insistence on certainty. If evidence only renders a conclusion *probable*, the cognizer must still rely to at least some degree on chance for being right. But then, as P. Klein (2000) pointed out, credit or blame is inappropriate if chance is involved and outcomes are not under the cognizer's control.

Normative Definition of Critical Thinking #6. Classical foundationalist

Purpose	*To accept only what is known with certainty.*
Constraints	*1. Only beliefs and inferences that are self-evident provide certainty.*
	2. Specific types of beliefs are intrinsically self-evident (e.g., those based directly on sensory appearances, introspection, or logic).
	3. Specific kinds of inferences are self-evident (i.e., those licensed by deductive logic).
Functions	*Critical thinking is:*
	(1) the identification of consciously available evidence for beliefs,
	(2) the independent evaluation of both the premises and the inferential steps in an argument, using (a) criteria of acceptance for premises that are satisfied by specific classers of intrinsically self-evident beliefs, and (b) criteria of logical validity of inferences that guarantee transmission of truth from premises to conclusion.
	(3) acceptance or rejection of beliefs based on the evaluation.

Implications for Dialogue

Figure 18 applies this definition to the evaluation of the belief labeled "A " in Figure 17. Figure 17 and Figure 18 illuminate three important implications of foundationalism for the role of *arguments* in reasoning and knowledge. Each point concerns the role of one of the three types of challenges illustrated by Types 1, 2, and 3:

1. *Modularity*. Evaluation of a belief p involves a dialogue like Type A, in which reasons for p are demanded. The chain of questions and responses (*Why <conclusion>?*) ends when it reaches basic beliefs at the bottom of the pyramid, where *Why?* questions are no longer appropriate. Thus, these challenges will never lead beyond the boundaries of a particular segment of the system of beliefs. As shown in Figure 17, only the beliefs *under* p in such a pyramid need be considered as parts of the argument for p. No other beliefs can be relevant.

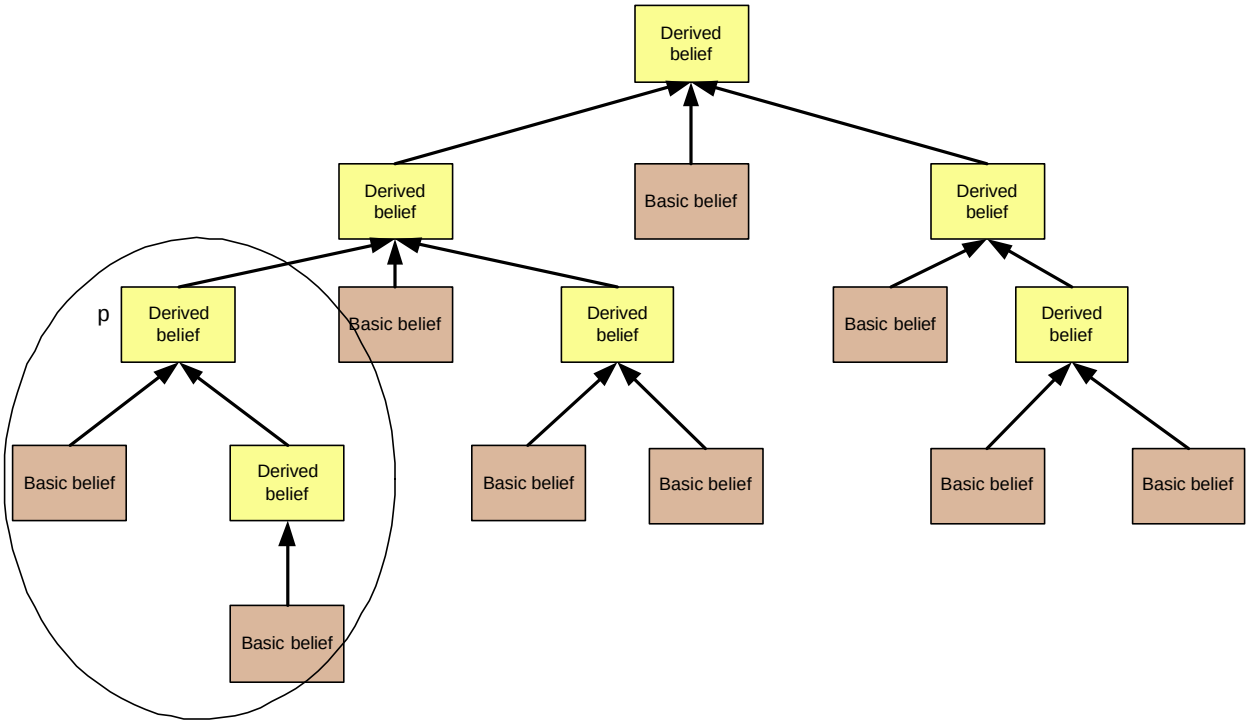


Figure 17. Modularity of a belief system due to presence of basic beliefs, where justification comes to an end.

2. *Independence of inference sufficiency and premise acceptability*. Once an argument is made explicit, Type B challenges (*Even if <reason>, possibly not <conclusion>*) are restricted to checking for logical validity. Such challenges may mention conditions under which the premises are true and the conclusion false, but they cannot introduce new information. They must involve logical combinations of propositions already introduced in the premises and conclusions as they stand. Moreover, because the criteria of premise acceptability and inference evaluation are independent of one another, evaluation of the argument for p can be broken down into two steps, as shown in Figure 18: (a) The *inference* of p from the reasons for p is evaluated in terms of logical validity (Type B). (b) The *reasons* are evaluated for acceptability (Type A). If a reason is a basic belief, it is automatically accepted. If a reason is not basic, we demand reasons for that reason (more Type A) and then iterate steps (a) and (b).

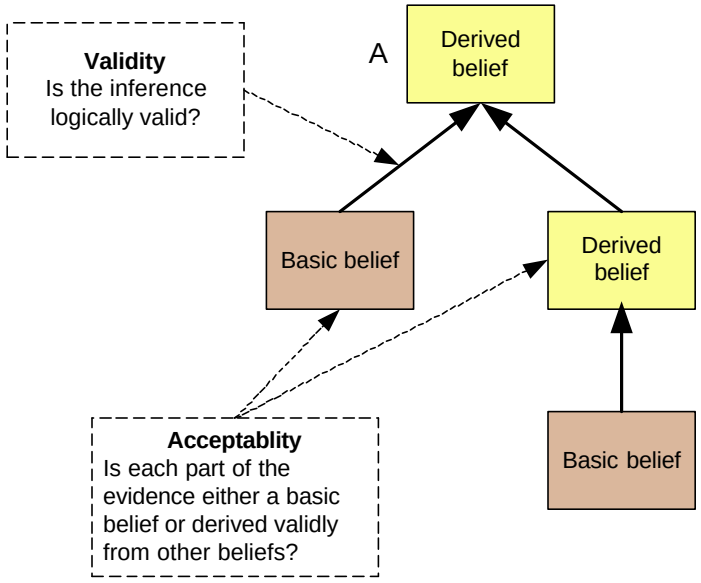


Figure 18. Evaluative criteria applied to an argument, according to the classical foundationalist paradigm.

3. *Uniqueness*. Challenges of the kind illustrated in Type C (*Not <conclusion>*) cannot occur at all. A logically valid argument from premises known to be true cannot be constructed for both p and not-p, assuming that the rules of logic are consistent. Thus, if the permissible Type A and Type B challenges have been answered, that is, if p is justified by a pyramid of reasons, it is a waste of time to look for arguments *against* p or to consider *alternative* views in which p is false. Once the truth of p has been assured, good arguments against p are impossible.

In sum, in classical foundationalism Type A challenges are strictly limited to basic beliefs. Type B challenges are limited to logical combinations of propositions already introduced by premises and conclusions. Type C challenges are not permitted at all. Argument is based on a modular part of the belief system, composed of independent steps, and immune to conflict. Foundationalism thus imposes a dialogue context in which there is a fixed starting point (basic beliefs) and every new assertion must be defended by tracing it back (via a chain of reasons) to that starting point. These features follow from two key principles: basic beliefs and uni-directional inference, each characterized by certainty. In this century, philosophers have successfully demolished both of these key components of the classical foundationalist paradigm.

Is Critical Thinking Loosely Based on Soft Foundations?

Contemporary foundationalism (Chisholm, 1977; Sosa, 1991) acknowledges uncertainty, both in premises and inferences. It thus rejects the two claims in classical foundationalism that are most problematic: (i) that basic beliefs are known with certainty, and (ii) that inferences guarantee transmission of truth from belief to belief. But contemporary foundationalism is still foundationalism. It has the same rationale as the classical version: The regress problem can be resolved only if there are two distinct classes of beliefs with inferences running in a single direction between them. Contemporary foundationalists claim that this solution of the regress problem still works even if beliefs and inferences are uncertain. Its success or failure hinges on

that claim. So in fact does much of the current work in informal logic and critical thinking, which are based on contemporary foundationalist ideas.

Because fallibility is acknowledged, the category of basic beliefs can be expanded beyond those recognized by classical foundationalism (i.e., beliefs reporting sensory experience, introspection, or logical truths). There are many more points at which the series of *why* questions can stop. For example, Plantinga (1993b, p. 183) argues that "many kinds of beliefs can be properly basic," including perceptual beliefs about physical objects, memory beliefs, beliefs based on the testimony of other people, and beliefs about the intentions or other mental states of other persons. Although the category of basic beliefs is now broader, it must still be restrictive. Many critical thinking textbooks state or imply that only certain *types* of beliefs are acceptable as ultimate premises in arguments. Dauer (1989; Chapter 2) lists the following types of typically "unproblematic" claims, which are in need of no further support by argument unless there are specific reasons to doubt them: observational claims, particular factual claims, intuitive claims, general claims of science and mathematics, and general claims of common sense. According to Dauer, claims in these categories, while not certain, are as sure as anything can be. That is, in the absence of *specific* defeaters (e.g., challenges against the veracity of a witness or the favorability of visual conditions), any doubts raised against basic beliefs would be so general that they could be raised against any claim whatsoever (p. 10). In short, the conception of *basic belief* held by contemporary foundationalists has two components: (i) Their justification does not require inference from other beliefs (Lehrer & Paxson, 2000, p. 31), but (ii) they are known as well as we know anything at all (Chisholm, 1977).[67]

Just as basic beliefs need not be known with certainty, it is also not necessary that the conclusion of an argument inherit *all* of the credibility of its premises. Descartes believed that all correct inferences were deductive, in which truth of the premises guarantees truth of the conclusions. But most of the inferences that occur in everyday tasks are *defeasible* (Pollock, 1995). That is, there are possible circumstances in which the premises are true and the conclusion false. Evidence that such a circumstance exists can be brought forward to *defeat* an inference, as a Type B challenge. The circumstances themselves are known as *defeaters*. This concession opens the door to a variety of non-deductive types of inference, such as enumerative induction, inference to the best explanation, and analogy. As shown in Figure 19, challenges to both premises and inferences (i.e., Type A and Type B challenges) are crucial to belief evaluation in contemporary foundationalism and informal logic.

[67] Contemporary foundationalists claim that this is all that is needed to stop the regress of reasons; it is not necessary that basic beliefs be *completely justified* or known with certainty (Bonjour, 2000).

Example of Type B Challenges

COL Black has faced a particular enemy commander, General X, on several occasions. On each of those occasions, General X moved his troops more rapidly than expected based on the usual formulas. COL Black predicts that General X will move the troops under his command more rapidly on this occasion as well. But speeding up her own operations will be costly in its effects on another important operation. So she gives the matter more thought.

She realizes that she made two inferences: that General X was responsible for the increased speed of his troops in the past, and that the present situation will resemble those past situations in relevant respects. So, she first asks herself: Could the past observations be true, but General X not have been responsible for the extra speed? She thinks of some possibilities. The speed observed previously might have been due to favorable terrain, new equipment, well trained troops, or able staff rather than to General X. Then she asks herself whether the present case might be relevantly different from the past situations (even if General X was responsible for the extra speed). She thinks of some possibilities: This time General X might have to wait for fuel supplies or synchronize with another unit. These are matters about which she can obtain further intelligence in order to gauge the risk and decide what steps to take to mitigate it.

Informal logicians agree that COL Black's reasoning involves an example of non-demonstrative or defeasible inference. Most, if not all, further agree that such inferences should not be "reconstructed" or interpreted to make them fit the deductive paradigm. Beyond this point, there is less consensus. There are a number of different ways that non-deductive inferences like the one in the example might be classified. Indeed, this ambiguity is a major practical problem in argument analysis:

(1) The inference might be considered *enumerative induction*. COL Black observes some cases and at first simply summarizes them:

> Summary. *Whenever I observed General X to be the enemy commander, I observed that enemy forces moved more rapidly than the formulas predict*.

She then inductively infers a generalization that extends to *all* instances:

> Generalization. *Whenever General X is the enemy commander, enemy forces move more rapidly than the formulas predict.*

She then deductively applies the generalization to predict a new case:

> Deduction. *General X is the enemy commander now. Therefore, enemy troops will move more rapidly than the formulas predict.*

(2) Another alternative is to take this as an example of *abduction*, or inference to the best explanation (Harman, 1986; Lycan, 1988). On this view, the summary (S) leads first to a causal hypothesis which is used to explain the past observations:

> Causal explanation. *General X's presence as commander is the best explanation of the fact that enemy forces under his command move faster than the formulas predict.*

A generalization (G) might be inferred from this causal hypothesis, and applied deductively (D) to predict the new instance, as in the previous example.

(3) A third way to construe the same reasoning is as an *analogy*, i.e., direct inference from previous cases to a new situation based on similarity in relevant respects. In reasoning by

analogy, there is no need to infer or explicitly formulate an intervening generalization or causal hypothesis:[68]

> Summary. *General X was present as commander at times $t_1 \ldots t_{n-1}$ and enemy forces under his command moved faster than the formulas predict at times $t_1 \ldots t_{n-1}$.*
>
> Analogy. *The present situation at time t_n is similar to the previous situations $t_1 \ldots t_{n-1}$ in that General X is present as commander.*
>
> *Therefore, enemy forces under his command will move faster than the formulas predict at time t_n.*

Regardless of how it is interpreted, this inference involves, either explicitly or implicitly, causal inference and causal knowledge. Similarly, whether it is construed as enumerative induction, analogy, or abductive / explanatory reasoning, it aims to make its conclusion probable, not certain.[69] Virtually every argument used in everyday life is *defeasible*. Classical foundationalism does not accommodate inferences that fail to guarantee the truth of their conclusions. It cannot handle the full array of Type B challenges.

Although many inferences are not deductively valid, contemporary foundationalists still regard inferential relations between evidence and conclusions as objective rather than subjective. Despite defeasibility, contemporary foundationalists need *objective* criteria of non-deductive validity in order to avoid skepticism. Critical thinking textbooks share this concern. Typically, they identify specific approved *types* of inferential transitions, such as deductive, enumerative induction, abduction, and analogy. If an argument satisfies the conditions associated with one of these specified forms of inference, its conclusion properly *fits* its evidence.

---

[68] An analogy might nonetheless use causal knowledge in determining similarity.

[69] Some authors use "induction" to refer to all probable, non-deductive reasoning.

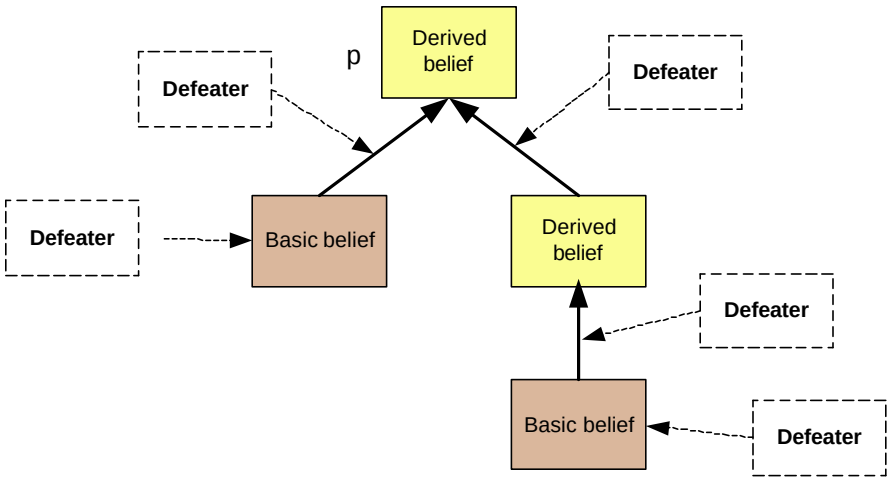


Figure 19. Defeasible inference and defeasible basic beliefs in contemporary foundationalism and informal logic.

Implications for Dialogue

The resulting picture differs from classical foundationalism only in the details: Reasoning constructs a pyramid, whose foundations are [a variety of different kinds of possibly uncertain] basic beliefs and whose higher levels are derived by [a variety of different kinds of possibly uncertain] inference rules applied to beliefs on the level below. Contemporary and classical foundationalism have similar implications for the role of argument with respect to modularity, independence, and uniqueness. Arguments are still modular, since the justification of a belief depends only on the beliefs under it in a pyramid. After eliciting reasons, we ask two kinds of questions: *Are the reasons acceptable?* as in Type A, and *Is the inference sufficient to support the conclusion?* as in Type B. Since one answer depends on criteria of sound inference and the other on identification of basic beliefs, the two steps are independent of one another. The one point on which contemporary foundationalism appears to diverge from classical is the possibility of conflicting arguments, as in Type C. Since both basic beliefs and inferences are fallible, it is possible to have reasonable arguments on both sides of an issue, for p and not-p.

This difference does not amount to much. It is quite surprising to discover that conflicting evidence plays a minimal role in contemporary foundationalism, informal logic, and critical thinking theory. Perhaps our analysis supplies an explanation for this neglect. Since all the relevant evidence must still appear in a modular part of the belief system (under the conclusion in a pyramid), pro and con evidence can always be combined, and the conclusion determined by evaluating a single, unique argument. By means of this device, informal logic texts generally manage to avoid the presentation and defense of separate positions, as in Type C challenges. There are no special strategies for handling conflict over and above assessing the acceptability of reasons (Type A) and the strength of inferences (Type B). It is at first surprising how little attention is paid by informal logicians to the resolution of conflicting arguments, as contrasted with the evaluation of individual arguments. But contemporary foundationalist assumptions provide the explanation.[70] Conflicting evidence is treated like a Type B challenge to the

[70] Govier (1987) is the exception that proves the rule. She proposes a *balance of considerations* argument type

sufficiency of an inference to the original conclusion. A conflicting piece of evidence may even be modeled as a defeating condition for a conclusion.

Contemporary foundationalism implies that the core critical thinking skills include recognizing arguments, identifying their components (e.g., premises, conclusions, and defeaters), and independently evaluating the premises and the inference from premises to conclusion. These ideas are also the most prominent features of mainstream textbooks in informal logic and critical thinking (e.g., Govier, 1997; Johnson & Blair, 1994; Freeman, 1993). Here is a normative definition of critical thinking that fits contemporary foundationist assumptions. It is a variant of the internalist definition given earlier:

Normative Definition of Critical Thinking #7. Contemporary Foundationalist

Purpose	*To increase the chance of accepting justified beliefs.*
Constraints	*1. Only premises that are initially highly probable and inferences that transmit probability of truth from premises to conclusions can increase the chance of accepting justified beliefs.*
	2. Certain types of beliefs are initially highly probable (e.g., those based on perception, memory, testimony, common sense, mathematics, or logic).
	3. Certain kinds of inferences transmit probability (e.g., deductive, inductive, or abductive inferences).
Functions	*Critical thinking is:*
	(1) the identification of consciously available evidence for beliefs,
	(2) the independent evaluation of both the premises and the inferential steps in an argument, using (a) criteria of acceptability for premises that involve specific classers of initially probable beliefs, and (b) criteria of sufficiency for inferences that involve satisfying the requirements associated with specific types of inferences that transmit probability from premises to conclusion, and
	(3) acceptance or rejection of beliefs based on the evaluation.

A crucial question for informal logic and critical thinking is whether foundationalism (in its contemporary form) can successfully accommodate uncertainty. The answer will in large part determine the viability of the approach adopted by most of the current work in critical thinking.

Foundationalism Inhibits Critical Dialogue

Internalist models of justification vary in the challenges they permit (Table 8) and thus in the prominence and importance of the role of the opponent or critic: Classical foundationalism admits only Type A challenges (*Why?*). If the proponent cannot justify each premise by a chain of argument rooted in basic beliefs, she must retract her conclusion. *Why?* questions, however, are not permitted with regard to basic beliefs, and all inferences are certain and thus not subject

---

alongside enumerative induction, abduction, analogy, and so on, each with its own procedures and criteria of soundness. A balance of consideration argument combines all the evidence in exactly the same way that other types of inference do.

to Type B challenges. Finally, properly conducted reasoning can never support conflicting positions, so Type C challenges are ruled out as well. Contemporary foundationalism raises the critic's stature somewhat. Basic beliefs are fallible but can only be challenged with specific defeating conditions. They cannot in general be challenged by asking for reasons. Contemporary foundationalism admits non-deductive inferences which are subject to Type B challenges *(Even if…still possible…)*, although deductive inferences are not. If sufficient doubt can be cast on the assumptions underlying an inference to the conclusion, then the conclusion must be retracted. Finally, coherentism is open to the widest range challenges. It allows Type A (*Why?)* questions to be posed regarding *any* belief and Type B questions to be asked regarding any inference. It also opens the door to conflict via Type C challenges (*Not...instead...*). Conclusions are retained or rejected based not on the acceptability of premises and the sufficiency of inferences, but on the overall plausibility of the bodies of beliefs that contain them.

Foundationalism runs into trouble in specifying both intrinsic and relational criteria:

1. There is no solid ground for the base of the pyramid, because virtually every belief depends on other beliefs for its justification. Type A (*Why?*) questions are appropriate with regard to any belief.
2. The inferential steps that add new beliefs to the pyramid are not infallible. Any inference can be confronted with Type B challenges, viz., defeating conditions under which the inference fails (*Even if…still possible…*).
3. Linear argumentation is insufficient for adding beliefs to the pyramid, because sound arguments may exist on both sides of an issue. Foundationalism offers no way to choose between alternatives in the case of Type C challenges (*Not…instead…*).

Each of these problems points, in different ways, to the same solution, i.e., *coherentism.* We will briefly focus on each point in turn. Coherentism corresponds to a dialogue in which the participants can begin from any mutually agreed upon starting point, need not defend or revise an assertion unless it is specifically challenged, and must defend or revise it if it is challenged. The critic not only has free reign to pose challenges where relevant, she may also present and defend a position of her own in opposition to the proponent's.

Table 8. Different types of challenges and constraints on challenges are associated with traditional epistemological theories.

	Type A *Why?*	**Type B** *Even if…still possible…*	**Type C** *Not … instead….*
Classical foundationalism (Internal analytic)	***May challenge only non-basic beliefs***	*Never*	*Never*
Contemporary foundationalism (Internal Empirical)	***May challenge only non-basic beliefs***	***May challenge only non-deductive inferences***	*Never*
Coherentism (Internal Empirical)	***May challenge any beliefs***	***May challenge any inference***	***May challenge any conclusion***

Any Belief Can Be Challenged

Type A challenges do not come to an end at self-evident beliefs. Although different beliefs vary in their degree of credibility, virtually any belief can turn out to be mistaken, even those that seem to directly report perceptual experience (e.g., Sellars, 1956/2000).[71]

[71] In addition to perceptual beliefs, so-called logical truths can also turn out to be false. Twentieth century mathematics (e.g., Russell, Gödel) is in part a response to paradoxes, which are examples of "self-evident" proofs of unacceptable conclusions. These conclusions motivate changes in the overall logical system or in our higher-order beliefs about it. Moreover, because of the role they play in scientific theories, logical beliefs are, like scientific hypotheses, subject to revision pressure when changes would better accommodate empirical data, e.g., in quantum physics (Everitt & Fisher, 1995; Quine, 1970).

Example

MAJ Jones believes that she saw a tank close by, out in the open, and in bright sunlight. A tank is an easily recognized object and visibility conditions are excellent. Nonetheless, this is not a good candidate for a basic belief, since beliefs about physical objects may turn out to be wrong (e.g., it could be a dummy tank or a mirage). MAJ Jones wishes to be a very careful critical thinker (in the classical foundationalist mode), so she asks *why* she believes there is a tank. To answer the question, she tries to focus on what she really "sees," that is, the tank-like shapes and colors in the visual image. Beliefs about these should be truly basic, she thinks.

Unfortunately, she runs into some problems. (i) She finds the task of focusing on raw "appearances" rather than the actual physical objects difficult and unnatural, as well as time consuming. (ii) Also, she has no convenient vocabulary to express these sensory thoughts about shapes and colors. (iii) And even if she could describe the bare appearances, she has no reliable inference rules that enable her to derive the existence of a tank from appearances. (iv) Even if she could describe the appearances and infer the existence of a tank from them, she finds that beliefs about experiences are not infallible after all. MAJ Jones takes another look and realizes that she underestimated the height of the turret-like shape relative to the rest of the image; perhaps her perception of the height of the turret was distorted by expectations built up in the past. She thus had formed mistaken beliefs even about these "self-evident" appearances! She gives up trying to obtain certainty, and reports with great confidence that she has seen a tank.

Contemporary foundationalism responds to these considerations by declaring all (or almost all) beliefs to be fallible. Nevertheless, it retains the idea of basic beliefs in order to resolve the regress problem. But if all beliefs are fallible, what makes some beliefs basic and others not? If there is no principled way to tell the difference, it makes no sense to give some beliefs a privileged status over others in dialogues like Type A (Lehrer, 2000, p. 82-83). It would lead to exchanges like the following:

> Proponent: *There is a tank.*
>
> Opponent: *Why do you believe there is tank?*
>
> Proponent: *I don't have to tell you why I believe it. Trust me, "There is a tank" is a basic belief.*

This kind of answer violates a fundamental rule of cooperative, rational dialogue, according to which a proponent of a position must not refuse to defend her position if challenged (van Eemeren & Grootendorst, 1992). Nonetheless, contemporary foundationalism provides a rationale for such a refusal in its definition of basic beliefs: They are (i) as justified as any belief can be and (ii) immune to support from other beliefs. For both these reasons, the question *Why do you believe A?* is said not to apply to them. Does this rationale hold water, or is the refusal to respond in fact a breach of reasoning protocol?

The problem is that basic beliefs are also supposed to be uncertain. But to say that beliefs are uncertain is to say that they are defeasible. That is, there are conditions under which a cognizer with such a belief might be mistaken. Basic beliefs are therefore only *prima facie* justified, subject to rebuttal by defeating conditions. Basic beliefs can be challenged by raising the possibility that a defeating condition is the case. And evidence that a defeater does not hold is a perfectly appropriate answer to the question *Why do you believe A?*

This raises a problem for the whole concept of basic belief. If we learn that a defeating condition about which there was some uncertainty is *not* the case, confidence in the basic belief will *increase*. Even if acceptance is all-or-nothing, our estimate of the chance that we will ever have to relinquish the belief is now less, since one possible circumstance in which that would happen has been eliminated. We now have two choices: First, we can admit that a "basic" belief can be supported by another belief. But this violates one of the two defining conditions of basic beliefs. The other alternative is to say that as soon as a belief gets support from another belief, it stops being basic. But then the property of being basic depends merely on whether the belief has in fact been questioned and support for it offered. Moreover, if a basic belief receives additional support from other beliefs, its credibility would increase even if it became non-basic. Thus, there would have to be non-basic beliefs with *more* credibility than basic beliefs. This violates the other essential property of basic beliefs, that they be as justified as any other beliefs are. Either way, the definition of basic beliefs in internally inconsistent, and the distinction between basic and derived beliefs is empty. [72]

Example

MAJ Jones believes that she saw an enemy tank at medium distance. Since a tank is an easily recognized object and visibility conditions are excellent, this is a good candidate for a basic belief (according to the more liberal standards of contemporary foundationalism).

MAJ Jones now learns that the enemy may have deployed dummy tanks in the region, since they have done so in similar operations in the past. This non-basic belief (an intel report about the possible presence of dummy tanks) is a defeater. It trumps her confidence in the *perceptual* judgment that what she sees is a tank. A basic belief can be undermined if it clashes with other beliefs which seem *less* secure.

But later, MAJ Jones learns that the enemy has not deployed dummy tanks in this region. She is now more confident that she saw a tank than she was at first. Now there is a dilemma: If her belief that she sees a tank is still "basic," we have a basic belief that depends on another belief, that the enemy has not deployed dummy tanks! If her belief is no longer basic, then we have a non-basic belief that has more credibility than the previous basic belief!

This is not to deny that some types of beliefs tend to be more trustworthy than others and that we generally expect some kinds of beliefs to be true in appropriate circumstances. For example, perceptual beliefs tend to be more reliable than beliefs based on reasoning or memory. But there are defeaters associated even with the most reliable beliefs. For example, perceptual

---

[72] We have argued that beliefs become more trustworthy when we learn that anomalous circumstances are not the case – just as the belief becomes *less* trustworthy if a cognizer learns that the anomalous circumstances *do* obtain. This symmetry follows automatically if degree of belief is represented as the relative proportion of possible situations in which a claim is true. In particular, in a Bayesian probability framework, if E is evidence against a proposition P, then it is mathematically necessary that not-E will be evidence in favor of P, though E and not-E need not have the same force. But the symmetry does not hold in some default logics (e.g., Reiter, 1980), which violate the logic of proportionality. They in effect treat acceptability as all-or-nothing, and claim that basic beliefs (unlike other beliefs) are accepted until we get information to the contrary. In this framework even when anomalous conditions are ruled out, it has been claimed that basic beliefs do not become *more* justified than they originally were. Even in this all-or-nothing framework, however, learning that a defeater is false surely decreases our estimate of the likelihood that we will ever have to give it up. This change can have practical effects. For example, it reduces our tendency to verify the correctness of that belief if a set of beliefs of which it is a member proves to be inconsistent..

beliefs are not trustworthy if viewing conditions are poor, vision is abnormal, the viewer is unfamiliar with the type of object being identified, or there are intentionally deceptive decoys in the vicinity (Lehrer, 2000, p. 72). In such situations, we may prefer memory or reasoning to the our own eyes. Similarly, testimonial evidence from others is trustworthy unless its source is not in a position to know, is dishonest, misspeaks, or is misunderstood.

In cooperative, rational dialogues, the proponent of a belief ordinarily has the burden of proof and is obliged to defend the belief whenever challenged to do so (Type A). Some beliefs, on the other hand, are accepted by default in certain circumstances – simply by virtue of someone's believing them. No clarity is gained, however, by regarding such default beliefs as fundamentally different from other beliefs. They are accepted by default for the simple reason that they are *reliable* in the relevant circumstances – that is, simply believing them is in these conditions a good indicator that they are true. For that reason, it makes sense that the burden of proof should be on the critic in those circumstances. Because of their high reliability, we demand more evidence of their falsehood before we question them. If we were to ask someone why she believes the testimony of her own senses, she would have a right to assume that we had some *specific* reason to doubt her perception in this case. Her response might even be *Why do you ask?* And she is obliged to defend her belief only if we respond to that question with a specific objection to it.

Default logics typically designate some beliefs as defaults, but the logic of such systems do not include any explanation of *why* those beliefs are selected and not others. To say that the default beliefs are intrinsically self-justified, i.e., basic, is both question-begging and incorrect. First, as we have seen, their default status does not prevent other (non-basic) beliefs from supplying part of their justification. If challenged as to whether she actually saw a tank, MAJ Jones can respond that viewing conditions were good, she has good vision, and she knows her tanks. None of these responses is itself basic or immune to challenge. Second, even if other beliefs were not relevant to their justification, a change in the conditions underlying their reliability would shift the burden of proof and rob them of their default status. It follows that default status cannot be due to any enduring properties of the beliefs themselves. Third, stakes play a role in where we locate the burden of proof. It is a waste of time to demand a defense of a reasonably reliable belief *unless the stakes are sufficiently high* to offset the reliability. If high stakes can rob a belief of its default status, then, once again, default status cannot be based on enduring properties of the beliefs. Indeed, a good case can be made that, when the stakes are low, we tend to give *all* our beliefs the benefit of the doubt. We retain *any* belief we actually form until we have specific reason to doubt it (Harman, 1973).To do so is perfectly reasonable, since computational limitations prevent us from starting from scratch and attempting to defend everything we think we know. Default logics, whatever their other merits, do not provide a rationale for an enduring basic-derived distinction.[73]

---

[73] Foundationalists' responses to their critics tend to trivialize the basic-derived distinction. One response is to drop the attempt to define *classes* of basic beliefs distinguished by intrinsic properties such as perceptual content, memory, introspection, or logic. Even though all beliefs are susceptible to support by other beliefs, it may the case that a particular belief is not *in fact* supported by any other beliefs (Audi, 1998, p. 207; for a related argument, see van Cleve, 2000; Allston, 2000). Thus, a basic belief is defined as a *particular* belief that just happens not to currently be based on any other beliefs. As soon as another belief is introduced as a reason for accepting it (e.g., about reliability of the current viewing conditions), the belief in question ceases to be basic. On this view, basicality is such a fleeting property that it tends to vanish as soon as we reflect on a belief at all! If the belief is challenged, or

Contemporary foundationalism and informal logic do not succeed in incorporating uncertainty into a pyramidal belief architecture. If basic beliefs are uncertain, they are defeasible. And if they are defeasible, then they can not only be defeated but also *supported* by other beliefs, which typically are not themselves basic. Thus, the basic-derived distinction fails. With it goes the notion that the segment of the belief system that supports a particular conclusion is *modular*, i.e., confined to beliefs *under* that belief in a pyramid. There is no clear point at which the series of *Why?* questions (and the sequence of *Even if…possible…* challenges) gets grounded at the "base" of a pyramid. Thus, challenges and responses can continue (at least in principle) until the entire belief system has been brought into the conversation. The attempt to stop an infinite regress of reasons has failed. The architecture of beliefs may better approximate a network (Figure 23) than a pyramid (Figure 16). And reasoning is better captured by dialogue rules that (i) permit *any* statement to be challenged and defended, but (ii) vary the threshold that a challenge must meet to shift the burden of proof to the proponent, and (iii) permit any statement to be accepted as long as it is *not* challenged.

Any Inference Can Be Challenged

There are still many adherents to the tradition that *all* correct inferences are deductive, i.e., the truth of the premises guarantees truth of the conclusions. Informal logic and contemporary foundationalism, however, have dissented from this tradition. One of the defining themes of informal logic is that most of the inferences that occur in everyday tasks are *not* deductive but rather *defeasible* (Pollock, 1995). There are possible circumstances in which the premises are true and the conclusion false. Evidence that such a circumstance exists can be brought forward to *defeat* an inference through a Type B challenge. Nevertheless, many informal logicians, contemporary foundationalists, psychologists, and artificial intelligence researchers still agree with the deductivist tradition that deductive inferences, *when they do occur*, are *not* defeasible. Deductive inference is therefore in a special class by itself – immune to Type B challenges just as basic beliefs were once thought to be immune to Type A challenges.

The restriction of Type B challenges to "non-deductive arguments" runs into problems. Consider the following:

> *MAJ Sud: Well, I don't agree that the enemy will attack in the north. They don't have any artillery over there.*
>
> *MAJ Nord: But don't we have reports that the enemy has developed longer-range artillery?*

---

if we subject it to critical thinking, the belief will instantly cease to be basic. But if basic beliefs can play no role in *reflective* reasoning, they cannot be foundational in any useful internalist sense. We can never *know* that a belief is basic, because as soon as we form a second-order belief about its reliability, it is no longer basic.

Another line of response is exemplified by Fumerton (2001) and Bonjour (2001), who defend a notion of direct acquaintance with experience (Fumerton) or intrinsically self-aware experiences (Bonjour). Neither of these manages to escape the original foundationalist dilemma, however. Either the entity that is basic is a belief or it is not a belief. If the basic entity is a belief, then there are conditions in which it can be incorrect, leading to the problems surveyed above (e.g., it can be supported by other beliefs that rule out those conditions, hence, it is not basic). If the basic entity is not a belief (e.g., it is an experience or act of apprehension), then it has no propositional content, which is to say that it cannot be true or false. Thus, it cannot be used in reasoning to narrow down the range of possibilities to a smaller set, and thus cannot logically justify beliefs. It may justify by virtue of being a *cause* of beliefs, but that implies an externalist view of justification, since causal relations are not directly accessible to consciousness.

MAJ Sud has just provided a brief argument that the enemy will not attack in the north based on their failure to place artillery in that area. MAJ Nord does not deny Sud's premise that there is no artillery in the north. Nor does he deny that this premise supports Sud's conclusion. What he does is offer a further bit of information that *neutralizes* the support given by Sud's evidence for Sud's conclusion. The absence of artillery in the north *in combination with* the fact that the enemy has developed longer range artillery is *not* evidence for the conclusion. Thus, MAJ Sud may have to retract his conclusion.[74]

Defeasibility in an open-ended, inescapable aspect of reasoning about the real world. Thus, the defeater itself is open to defeat by additional information. Consider MAJ Sud's reply:

> *Officer Sud: True, but I don't recall any indications that they've deployed the new systems yet.*

MAJ Sud has presented a defeater for MAJ Nord's defeater. If they enemy has not deployed the new artillery, then possession of the longer-range technology is irrelevant to the location of the attack. The three claims in combination – lack of artillery in the north, possession of longer range artillery, and no deployment of the longer range artillery – do provide evidence for attack in the north. Thus, the force of Sud's argument is restored.

Defeasibility has been the principle reason for the rejection of formal logic as an overall framework. But formal logic might still be a useful tool for analyzing *some* arguments. If so, we need a way to determine which arguments it is useful for. Are there unambiguous descriptive criteria that a deductive inference has occurred or is intended? Reconstructing the intended argument from actual utterances is a major problem. Any argument, including the one above, can be reconstructed in syntactically correct deductive form by supplying "implicit" premises (we can simply add a conditional with the evidence of the argument as antecedent and the conclusion of the argument as the consequent). But let us by-pass that problem, because it is not clear that correct syntactic form is sufficient to identify a deductive inference. Suppose that MAJ Sud actually presented her argument in a syntactical form that is explicitly deductive:

> *Sud-1: The enemy does not have artillery in the northern sector.*
>
> *Sud-2: If the enemy does not have artillery in a sector, they will not attack in that sector.*
>
> *Sud-3: Therefore, the enemy will not attack in the northern sector.*

If this argument is deductive, then it is deductively valid. But is it deductive?. We saw that it was defeasible, and since defeasible arguments do not guarantee their conclusions, they are not deductive. Syntax is not sufficient to establish the presence of a deductive argument. But suppose Nord had *not* challenged the argument with a defeater, but instead had accepted it. Would it then have been legitimately regarded as deductive? But how could we or MAJ Sud know ahead of time whether a Type B challenge would be forthcoming? What if the argument was only challenged the next day, or the next week? Would it be deductive for a day or a week and then

---

[74] Artillery would normally be used to soften the opposing front line before an assault, so its absence suggests that no such assault is planned. The defeater states that longer range artillery could help soften the opposing front line in the north without being present there.

More technically, even if evidence E justifies a conclusion C, there may be other information D such that *E and D in combination* do *not* justify C. Then, D is a defeater for E in that context (Pollock & Cruz, 1999, p. 37). A defeater does not have to be evidence *against* the conclusion or *for* any another conclusion (although it might be). It may simply neutralize or cancel out the evidence for the conclusion.

suddenly and retroactively become defeasible? Does it matter whether or not Nord *intended* that defeaters be relevant? Does it matter whether Sud understood defeaters to be relevant? There is no syntactic or semantic mark that enables us to tell ahead of time whether this argument is deductively valid (hence not defeasible) or defeasible (hence not deductive), and it is not clear what difference the intentions of the parties make.

Proponents of formal deductive models have tried various devices for promoting co-existence between deduction and defeasibility. The simplest tactic is to regard defeasibility as merely a problem with the *premises* of a changing, deductively valid argument. Let's see how that might work. In our example, Nord has challenged Sud-2. This is equivalent to proposing a new premise:

> *Sud-2¢: If the enemy does not have artillery in a sector, and if they have not developed long-range artillery, they will not attack in that sector.*

Nord argues that the problem is more than hypothetical; there is reason to believe that the antecedent is true:

> *Nord: The enemy has developed long range artillery.*

MAJ Sud's argument is defeated, since substituting Sud-2′ for Sud-2 results in an invalid deduction. This defeat is subject, of course, to further rebuttal by Sud. Sud's response is equivalent to proposing yet another premise in place of Sud-2′:

> *Sud-2¢¢: If the enemy does not have artillery in a sector, and if they have not developed long-range artillery, and if they have not deployed long-range artillery, they will not attack in that sector.*

Sud also claims that the enemy has not in fact deployed such new artillery:

> *Sud-4: The enemy has not deployed long range artillery*

Thus, we interpret the dialogue as an iterative process of challenging and revising the premises of a deductive argument. The dialogue between Sud and Nord addresses the plausibility of the premises, while each member of the series of arguments within that dialogue is intended to be deductively valid.

There is a problem with this reconciliation between deductive reasoning and defeasibility if the revised premises are meant to become parts of the participants' belief systems. On the one hand, if Sud-2″ supplants Sud-2 in MAJ Sud's belief system, then she will be unable to reason with incomplete information in future situations. But in many circumstances where artillery is used as an indicator, there is no reason for MAJ Sud to even consider the development or deployment of longer range artillery. Artillery location alone is often a plausible indicator of an enemy's planned location of attack, as indicated by the original premise Sud-2.[75] On the other

---

[75] To make matters worse, the participants might be blocked from making a decision in the present situation as well. The list of potential defeaters is indefinitely long, and advance specification of all defeaters is probably impossible in principle. (The set of defeaters for the inference from an effect to a cause, for example, must include *all* the other possible explanations of the effect.) As the conversation between MAJ Nord and MAJ Sud continues, more exceptions and exceptions to exceptions may always be brought forward. Each new complication of the rule would cause a revision in the beliefs that serve as premises and thus ratchet up the demand for information before the inference can be regarded as valid. If the decision maker agrees that a defeater is relevant, she must add it to the antecedent of the premise. But if no information is available to decide its truth, she cannot make a decision regarding

hand, if *all* versions of the premise (Sud-2, Sud-2′, and Sud-2″) are retained in Sud's belief system, the decision maker will be able to reach conclusions based on incomplete information. However, none of the inferences will be defeasible. Sud-2 alone will always yield a valid inference, even if the more stringent conditions imposed by Sud-2″ are *known* to be false. In sum, construing this argument as deductive in either of these ways robs decision makers of flexibility to adapt the reasoning process to specific circumstances. They will not be able to act decisively on a subset of the relevant information in situations where that is appropriate (i.e., time is limited and/or stakes are low) but *also* spend more time thinking and gathering information where *that* was appropriate (e.g., time is available and/or stakes are high).

The stakes of the decision would influence the burden of proof. If the stakes are high, the proponent will have the burden of showing that a potential defeating condition does not obtain. If the stakes are low, the opponent might have the burden of showing that the potential defeating condition does obtain. But if stakes can affect the burden of proof in this way, then it is a mistake to treat defeating conditions as antecedents of conditionals in a deductive proof. To make sense of both defeasibility and reasoning with incomplete information, we need to discard the attempt to construe arguments like the ones above as deductive. When the conditional of the inference is made explicit (as in Sud-2, Sud-2′, and Sud-2″), it may be literally true that the premises guarantee the conclusion, and therefore the arguments are deductive. But the main focus of the dialogue is not on determining deductive validity, but on challenging the conditional premises and revising them to include new defeaters. To reconcile reasoning with incomplete information and defeasibility, we need dialogue rules similar to those we proposed for basic beliefs. *Any* inference (even if it looks deductive from a syntactic point of view) is subject to challenge by the opponent in a critical dialogue. The threshold of seriousness a challenge must clear in order to shift the burden of proof to the proponent will vary with the stakes,. That is, the cost of an error will determine whether the defeater must be shown to be true (low stakes) or must only be shown to be possible (high stakes) by the opponent. If the inference is challenged by a defeater that clears the threshold, the proponent must defend her conclusion against it. If it is not challenged, no defense is necessary.

The notion of deductive inference is not useful in the context illustrated above. It does make sense, however, in some specialized contexts. In such contexts, some inferences and premises are temporarily not subject to challenge. A deductive argument is a type of dialogue whose purpose is to draw out the logical implications of a fixed set of premises. A deductive reasoning dialogue (see *rigorous persuasion dialogue* described in Walton & Krabbe, 1995; also Hintikka, 1999) is likely to be embedded in a larger dialogue context in which a wider range of challenges are permitted. In such a context, it may sometimes be useful to suspend freedom of discussion temporarily in order to rigorously examine the implications of commitments already made by one or the other of the parties. In that case, the parties might agree (implicitly) to conduct a sub-dialogue in which the rules temporarily exclude direct challenges to the premises (Type A) and limit Type B challenges to those that concern logical validity of the inference. In particular, "defeaters" of the inference may be introduced only if they involve combinations of propositions already present in the premises, not if they introduce new events as defeaters. That is, they must introduce possibilities in which the premises (as they stand) are true and the conclusion is false. When the work of this sub-dialogue is done, the participants will resume the

---

the conclusion.

larger dialogue. A participant is then free to use the conclusion of the sub-dialogue in further reasoning, or alternatively to revise or reject one or more of the premises.

Any Conclusion Can Be Challenged By an Alternative Account

Following in the footsteps of foundationalists, informal logic texts depict argument evaluation as a process consisting of two independent steps: "There are two essential aspects of good arguments: (i) acceptable premises and (ii) a conclusion that follows from these premises" (Groarke, Tindale, & Fisher, 1996). Similarly, according to Govier (1997: p. 74), "The basic elements of a cogent argument …are as follows: 1. Its premises are all acceptable…2. Its premises are properly connected to its conclusion…" The idea that premises and inferences can be evaluated independently of one another is very important. This is what makes it possible to add beliefs to the top of the pyramid one argument at a time, each new inferential step building on the beliefs laid down previously. Without such independence, progress would no longer be cumulative; each new inference would demand a re-evaluation of the premises, and thus reopening previous arguments leading up to the new step.

Cumulative acceptance of new beliefs fails when there are conflicting arguments, i.e., challenges of Type C. Take the simplest possible case, in which we have one uncertain argument for p and another for not-p, as shown in Figure 20. Given uncertainty, it is possible that *both* of these arguments are sound *if taken separately* – that is, their premises are acceptable and sufficient to establish their respective conclusions. Since it is impossible for both p and not-p to be true, this approach is flawed; independent evaluation of each argument is inadequate. Some informal logicians who have recognized this problem recommend combining conflicting evidence into a single argument (e.g., Govier, 1987; Thomas, 1997). The combined evidence is then evaluated the same way non-conflicting evidence would be evaluated, in terms of premise acceptability and sufficiency. However, there is a problem here as well. It is very unlikely that the resulting combined argument will be sufficient to support *either* conclusion. The problem is two-fold: (i) We have stipulated that the premises are acceptable. (ii) Assessments of the *combined* evidence will never resolve conflict because it will never be "sufficient," as it stands, to establish either conclusion. The problem is the presence of strong evidence pointing in both directions among the premises of the combined argument.

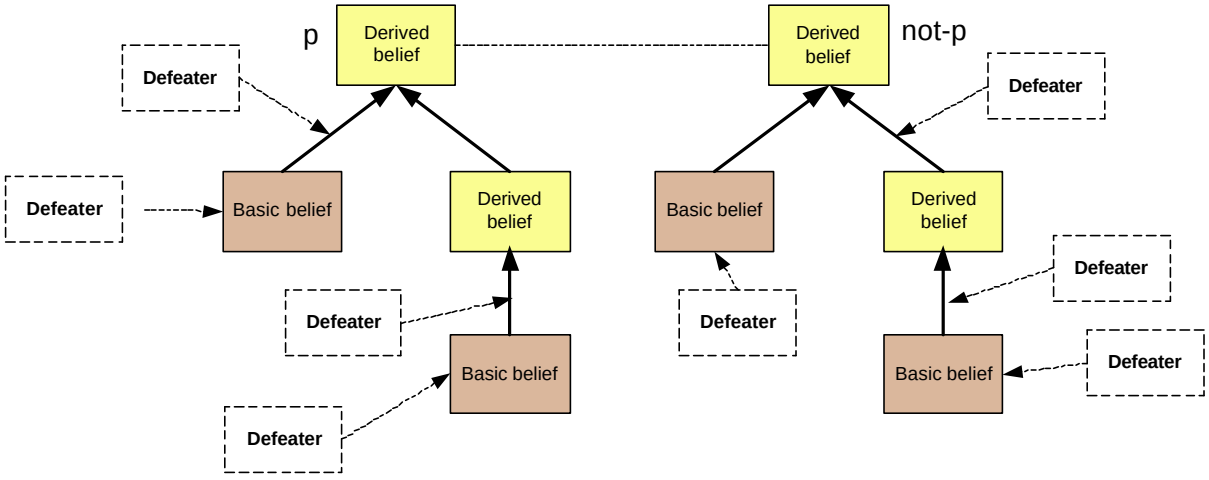


Figure 20. Two arguments support conflicting conclusions. (A dotted line between derived beliefs shows that truth of both together is either logically impossible or highly improbable.)

What is missing is some way of *explaining* and thus neutralizing the evidence in one of the conflicting arguments. Conflict between two arguments is evidence that something is amiss in the beliefs that led to the conflict, but does not tell us what. There are many possibilities: The evidence for p is unacceptable *or* the inference from that evidence to p is insufficient *or* the evidence for not-p is unacceptable *or* the inference from that evidence to not-p is insufficient. (Each of these possibilities can be further broken down into separate pieces of evidence and separate inferential steps.) It is already clear that independence of premise and inference evaluation fails. We need to be able to re-open premise evaluation when an inference leads to an implausible result, i.e., one that conflicts with other evidence or reasoning. In the context of conflict, if we regard the *inference* for p as normatively adequate (i.e., "sufficient"), then more suspicion falls on the acceptability of the *premises* for p, and vice versa. Similarly, if we regard the *inference* for not-p as sufficient, then more suspicion falls on the acceptability of the *premises* for not-p, and vice versa. Premises and inferences cannot be evaluated independently in the presence of conflicting conclusions, but informal logic and contemporary foundationalism provide no rationale for this. On the contrary, they explicitly assert that premise evaluation and inference evaluation *are* independent.[76]

Suppose a cognizer's initial mental model of the situation contains the following three beliefs:

not-q p p → q

Logic tells the cognizer that this mental model is inconsistent and therefore cannot be true; hence, one or more beliefs in this set must be revised. But there is more than one set of changes that will work. For example, one way to restore consistency is to reject not-q and replace it with q. This corresponds to the following argument:

p: The enemy is concentrating artillery in sector S.

[76] A premise may be rejected if a specific defeater associated with *that* premise is subsequently found to be true. But the rejection is not based on the implausibility of inferences from the premise, and does not require search for the best solution.

p → q: If the enemy concentrates artillery in sector S, they are planning to attack in sector S.

These two premises are both accepted by the cognizer, and together they logically imply the following conclusion:

q: The enemy is planning to attack in sector S.

But the cognizer can rationally reject this conclusion! The cognizer need not accept q even though she believes both p and p → q. In this simple example, where three beliefs are inconsistent with one another, there are three direct ways to restore consistency, each of which revises just one of the conflicting beliefs. Moreover, a logically valid argument can be constructed for each of these revisions, based on the two beliefs that the cognizer continues to accept. Thus, any of the three consistent mental models shown in Table 9 can be "justified" by a logically valid argument that starts from premises the cognizer accepts. Since all three arguments cannot be endorsed, it is clear that logical entailment is insufficient for justification, and that there is no purely logical argument that distinguishes one consistent set of beliefs form the others. [77]

Table 9. Three logically consistent models.

Logically consistent models of the situation that involve a change in only one belief			**An argument (based on logical implication) leading from parts of the current model to a new model** Premises Conclusion	
q	p	p → q	p p → q	q
not-q	not-p	p → q	p → q not-q	not-p
not-q	p	not- (p → q)	p not-q	not- (p → q)

The inability of linear argument to resolve conflicting evidence is not confined to *deductive* arguments. The same point applies to arguments that establish the plausibility rather than the certainty of their conclusions given the evidence (i.e., defeasible inference as discussed by contemporary foundationalists and informal logicians). Any deductive or non-deductive argument can be taken to assert the "inconsistency" of the conjunction of its premises and the

[77] Harman (1986, pp. 15-16) and Lycan (1996, p. 10) argue further that logical consistency is not *necessary* for coherence. If we see no way to resolve an inconsistency between two sets of beliefs, or cannot resolve it without great effort, it may be rational to continue using each set of beliefs in its own sphere of application. Compartmentalization of beliefs, i.e., modularity, makes it unlikely that a contradiction will ever actually be inferred.

negation of its conclusion. If the argument is warranted by some principle other than deductive validity, then so is the corresponding notion of inconsistency. Thus, *non-deductive* concepts of consistency also fail to provide sufficient criteria of coherence. In cases of conflicting evidence, if there are multiple sound arguments for conflicting conclusions, there are also multiple consistent sets of beliefs, each involving revision of some beliefs on the basis of others that are not revised. Thus, if coherence is defined as consistency of any kind, it provides no response to the problem of resolving conflict.

It is worth noting in particular that the same problem arises for another proposed criterion of coherence, *probabilistic consistency* (Harman, 1999a). Uncertainty is a central feature of reasoning in many real-world contexts – including, of course, military tactics – where the relations among events are not known with certainty, either because they are not deterministic or because information is incomplete. Bayesian probability theory provides a way to quantify degrees of belief for propositions and to use the quantities that are known to derive degrees of belief that are unknown. The probability calculus can be used to define constraints that probabilistic judgments must satisfy on pain of inconsistency. But it is no more a logic of inference than deductive logic was. A choice must still be made among diverse ways of restoring consistency in cases of conflict. In fact, with probabilistic systems of beliefs, it is not necessary, and indeed implausible, to assign all the blame for inconsistency to just one of a conflicting set of beliefs. It is often more sensible to tune all the probabilities simultaneously to make them consonant with one anther. With this kind of latitude, there will be an unlimited number of consistent solutions. Thus, neither deductive nor probabilistic arguments directly justify *individual* conclusions. Rather, they help the cognizer evaluate the *global* coherence of the belief system and see how a particular belief is linked probabilistically or logically to other beliefs in the system. [78]

Another argument-centered approach is to construct arguments for all the different possible belief revisions and compare their strengths. Figure 21 gives an example of a series of arguments, each of which is sound from the perspective of informal logic. If taken alone, the evidence in the first of the two original conflicting arguments is both acceptable and sufficient to justify p (although not to establish it with certainty). Now we look at the second argument, not

---

[78] Suppose the cognizer begins with the following probabilistic beliefs:

Prob(q | p) = .9 Prob(q | not-p) = .2 Prob(p) = .7 Prob(q) = .35

Probability theory tells us that these beliefs in combination are inconsistent. One way to restore consistency is to revise Prob(q) from .35 to .69. This corresponds to the following argument or calculation, taking the other three probability assessments as given:

Prob(q) = Prob(p) Prob(q | p) + (1 - Prob(p)) Prob(q | not-p)

= (.7) (.9) + (.3) (.2) = .69

However, we are not compelled to accept this argument – even though that conclusion is validly derived from premises we accept. The probabilistic argument merely states a formal relationship, just as the logical arguments we looked at before; it is not the same as an inference that accepts a probabilistic conclusion. We might choose to retain our belief that Prob(q) = .35 and restore consistency by changing one of the other beliefs. For example, we could revise our estimate of the probability of p from .69 to .21. In that case, we could endorse the following argument:

Prob(p) = [Prob(q) - Prob(q | not-p)] / [Prob(q | p) - Prob(q | not-p)]

= (.35 - .2) / (.9-.2) = .21

This argument is also probabilistically valid and based on premises we accept.

Note that the same problems arise for non-Bayesian approaches to probability (e.g, when probabilities are interpreted as relative frequencies) – as long as it is possible for different measurements or assessment techniques to yield discrepant answers.

by itself but in the light of the results of the first argument. We can use the initial conclusion, p, as evidence that the second argument (which supported not-p) is unsound, i.e., there is a problem with one of its premises or the inference from those premises. If we still regard the premises of the second argument as acceptable, those premises, in conjunction with p, show that the *inference* from those premises in the second argument is faulty. We can thus conclude with confidence that at least one defeating condition for that inference is true (and is not itself defeated).

This is entirely reasonable within the framework of informal logic, which frames rational thinking in terms of sound argument. A critical thinker who reasoned this way could not be faulted within informal logic. But the pitfalls of this approach are clear: The result depends on which of the two original arguments the thinker chose to start with. Suppose she had begun with the second argument, the one with not-p as a conclusion, as shown in Figure 22. By exactly analogous steps, she could infer not-p, then use that conclusion to argue that a defeating condition in the argument for p is true and undefeated. The result also depended on other choices, e.g., her decision not to question the acceptability of the premises in the second argument, and thus to conclude that the inference of not-p was flawed. She might instead have used the sufficiency of the inference for not-p to conclude that one of the premises was unacceptable. She would then look for defeating conditions that apply to the premises.

The problem is that an argument-by-argument approach as promoted in informal logic provides no guidance as to how such decisions should be made, i.e., no mechanism for *coordinating* the results of individual arguments in a reasonable way. Using a series of arguments in this way might be appropriately dubbed the *argument fallacy*. Some arguments, taken in isolation, may appear to be cogent and thus warrant the acceptance of their conclusions, while other arguments do not. But because of the fallibility of each inference, the final result of accepting and rejecting conclusions *one at a time* may be an implausible overall model (e.g., acceptance of both p and not-p) or an arbitrary one (e.g., selecting p or not-p based on the choice of a starting point). This blatant path dependence is the fatal flaw in the argument-centered approach and is responsible for the argument fallacy.[79]

There are some important morals of this example. When there are conflicting opinions, a cognizer must never regard a single argument as the last word, *even if it includes all the available information* and *passes all the ARG criteria*. The arguments illustrated above cannot solve the problem either individually or jointly. If the cognizer considers only one of them, she runs the risk of dropping a belief that should be kept, or of retaining a belief should be dropped. If she considers both of the arguments, she may either continue to hold an incoherent set of beliefs or adopt an overall view that is implausible (by revising more of her beliefs than is necessary). Moreover the arguments cannot be diagrammed as parts of a single converging argument, as suggested by Thomas and Govier for conflicting arguments, since they do not pertain to the same hypothesis.

Ultimately, the problem with arguments for individual hypotheses is due to *defeasibility* of inferences, i.e., the possibility of encountering new information that forces the retraction of previous conclusions. But more precisely, it is due to the symmetrical roles that alternative explanations play as defeaters for one another. Thus, the engineering staff's report is unreliable

---

[79] Path dependence of this kind occurs in the *confirmation bias* (Nisbett & Ross, 1980), i.e., a tendency to ignore or discount evidence that conflicts with an initial hypothesis.

unless the G-2 staff's report was unreliable. And the G-2 staff's report is unreliable *unless the engineering staff's report was unreliable*. We could have expanded this example so that B and C explored additional possible explanations of the failed predictions. If we had, each of the explanations would be a defeater for all the others. Because of this symmetry, adding defeaters to the architecture of arguments does not make arguments for individual hypotheses more useful in the resolution of conflicts. A dialogue of warring arguments will go on tit-for-tat but will not resolve conflict unless the participants are able to assemble the implications of the arguments into a larger picture.

Arguments in informal logic lead to acceptance or rejection of individual claims. No method is provided for evaluating *as a whole* the set of beliefs that results from a series of arguments, or for revising earlier conclusions based on later ones. The ARG method assumes that locally optimal decisions with respect to each intermediate conclusion will yield a globally optimal system of beliefs, i.e., an adequate overall picture of the situation. But this is not the case when different arguments point to different conclusions. Failure to consider an ensemble of interrelated beliefs as a whole can lead to impossible or implausible models of the situation, and thus to a complacency that is incompatible with the goals of critical thinking. (Similar problems arise in picking a stock portfolio stock by stock rather than considering how they relate to one another to affect overall performance.) [80]

The role of arguments is in part to probe for problems in mental models, such as incoherence. It was the conflict between the conclusions of two arguments that first told B that his beliefs about this situation were flawed. But arguments for individual hypotheses cannot generally *resolve* differences of opinion. If the cognizer does rely on such arguments, she should construct a separate one for rejecting each of the alleged culprits, compare the force of those arguments in a way that aggregates premise acceptability and inference strength, and then use the results to build a coherent overall account that involves as few changes as possible from her original view. But this is equivalent to selecting the mental model with the highest probability. In this example that strategy can be implemented by revising the weakest element in the original model (i.e., the belief in the G-2's reliability).

[80] Pearl (1989) makes a similar point in the context of a probabilistic framework: "…by belief commitment we mean the categorical but tentative acceptance of a subset of hypotheses that together constitute the most satisfactory explanation of the evidence at hand. In probabilistic terms, that task amounts to finding the most probable instantiation of all hypothesis variables, given the observed data.[p. 240] …this optimal assignment cannot be obtained simply by optimizing the belief distributions of the individual variables [p. 246] ."

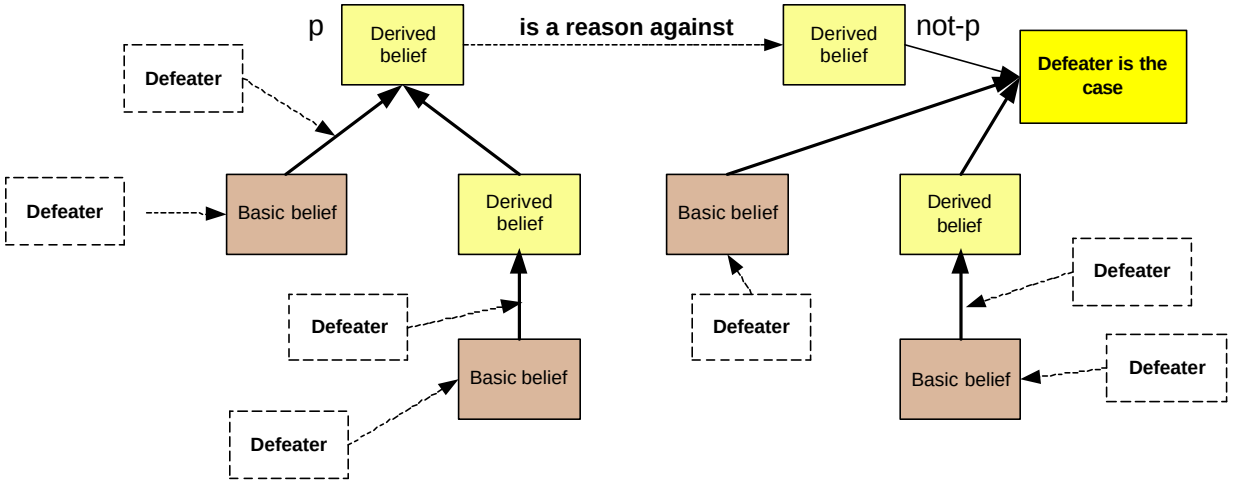


Figure 21. Once a conclusion (p) is accepted on the basis of argument on the left, it can be used as evidence against conclusion (not-p) of argument on the right. Falsity of that conclusion can be used as evidence that a defeater is the case.

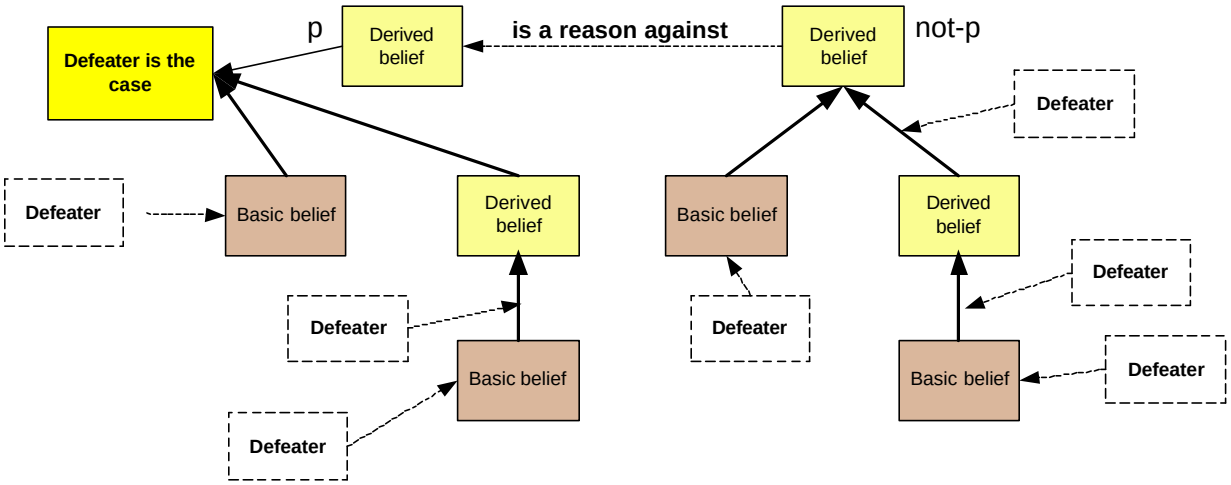


Figure 22. The entire process shown above can be reversed, using the conclusion (not-p) of the argument on the right as a reason against the conclusion (p) on the left, then using the negation of the conclusion as support for a different defeater.

In sum, foundationalism tries to limit the dependence of a belief on other beliefs. If the belief is basic, it depends on *no* other beliefs, and if it is non-basic, it depends only on the beliefs immediately below it in a pyramid of beliefs. But neither of these limits holds. By admitting defeasibility, contemporary foundationalism is able to acknowledge uncertainty and to handle Type B challenges. The price to be paid is that defeasibility undermines the concept of basic beliefs. Moreover, the *pervasiveness* of defeasibility undermines the linear, argument-by-argument derivation of non-basic beliefs under Type C challenges. When there is conflict, at least one defeating condition in one of the competing arguments must actually be true, but there

is almost always more than one *possible* defeating condition – and we cannot trust any individual argument or series of arguments to tell us which one is at fault. These two problems – the untenability of basic beliefs and the inability of linear derivation to resolve conflict – are distinct from another. However, they both arise from defeasibility and they both point to the same conclusion, that the credibility of any belief or inference can be affected (at least in principle) by virtually any other belief. Inference may ultimately depend on the evaluation of overall *systems* or *collections* of beliefs, rather than on relatively myopic arguments that lead from one belief to another.

9. WHEN IS A STORY COHERENT?

As a result of the problems with foundationalism, many philosophers have taken up a more sophisticated variant of the "circular reasoning" option called *coherentism* (Thagard, 2000; Lehrer, 2000; Everitt & Fisher, 1995; Harman, 1986; Bonjour, 1985; Quine & Ullian, 1970). From the coherentist perspective, there are no privileged beliefs that serve as foundations. Nevertheless, justification does not involve circular reasoning because it is the *system* of beliefs that is the primary target of justification rather than the individual beliefs within it (Day, 1989; Plantinga, 1993a, pp. 78-80). The architecture of a belief system is not a pyramid but a network, as shown in Figure 23. A system of beliefs is coherent and therefore justified when its members are tightly interconnected by explanatory, logical, conceptual, or other relationships. Every belief – perceptual, logical, scientific, or introspective – potentially contributes some support to every other belief and in turn draws support from every other belief. It takes a set of mutually supporting beliefs to generate a prediction, and when surprises occur, it is necessary to look at the whole set of beliefs, not just one, in order to find the most likely problem. Resolution of conflict requires evaluation and comparison of alternative mental models, not arguments for individual beliefs.

Example

Suppose MAJ Jones believes that she saw a tank. Since a tank is an easily recognized object and visibility conditions are excellent, this is a good candidate for a basic belief. But it can be undermined if it turns out to clash with other beliefs which on the face of it seem less secure. Suppose MAJ Jones learns that the enemy has deployed dummy tanks in the region, or remembers that the area where she "saw" the tank is shown as a swamp on the map. These non-basic beliefs may trump her confidence in the perceptual judgment! Alternatively, the perceptual judgment might lead MAJ Jones to question the map or the reports of dummy tanks.

MAJ Jones must determine which *overall set of beliefs* is most plausible, including beliefs about the presence of the tank, the accuracy of the map, the reliability of the reports about dummy tanks, and the reliability of his own perceptual judgment. In other words, MAJ Jones must evaluate the plausibility of alternative mental models. The decision whether there is a tank will depend on general beliefs about the accuracy of maps, intel reports, and perceptual experiences, which in turn depend in part on the past performance of similar maps, reports, and perceptions. That is, the selection of a plausible mental model will depend on its coherence with a larger body of beliefs. Each belief is justified by its coherence with the others.

A central problem of critical thinking is how to avoid an infinite regress of arguments – in short, to know when to *stop* demanding reasons for a belief, reasons for the reasons, and so on. Some possible answers are:

- Skepticism: Never – justification is illusory.
- Relativism: At assumptions that cannot themselves be justified.
- Foundationalism: At a rock-bottom set of beliefs, justified by their intrinsic properties (such as perceptual, logical, or introspective content) rather than by inference from other beliefs.

Coherentism's answer to the regress problem is more subtle. In the simplest cases reasoning in support of a new conclusion stops when it reaches *already accepted* members of a coherent system of beliefs. This kind of reasoning resembles the argumentation described by contemporary foundationalist, in which already accepted beliefs serve as reasons in arguments for the new belief. A difference is that they are not "basic" or privileged in any sense other than being accepted. They are not arbitrary assumptions either, since they are justified as part of a coherent overall system of beliefs.

Coherentism looks quite different when new information *conflicts* with already accepted beliefs (as in Type C challenges). In this case, to incorporate the new information into the system of beliefs, the cognizer evaluates competing bodies of beliefs in terms of criteria of coherence. It will be necessary to *revise* some of the already accepted beliefs (Gardenfors, 1992), and in such a revision process, virtually the entire network of beliefs may in principle be affected.

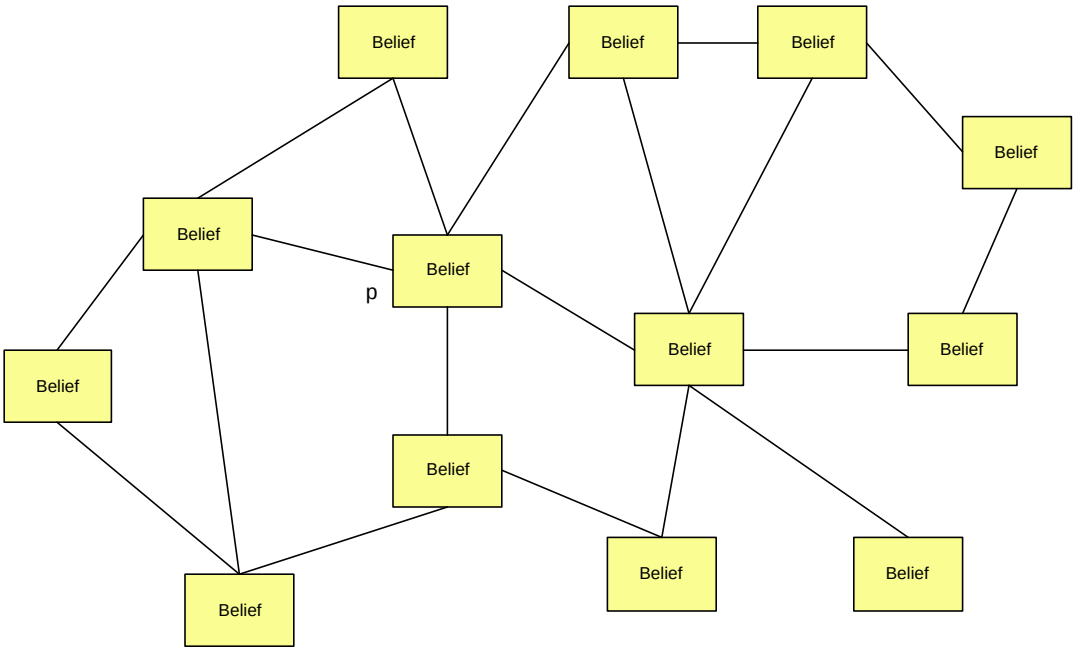


Figure 23. Coherentist paradigm for acceptability of beliefs: a network. The system of beliefs is justified as a whole by the inferential links among its components and its overall simplicity and comprehensiveness. Beliefs are not classified into types with different epistemological status, such as basic or not basic.

Arguments bear on justification *indirectly*, by exposing inferential relationships that contribute to the coherence of the system of beliefs as a whole. An individual belief is justified indirectly by having a place in such a coherent system of beliefs. Even perceptual beliefs, which were not acquired by inference from other beliefs, are justified in part because of other beliefs, e.g., about the reliability of visual processes under good conditions of visibility. Arguments are essential tools, since they may be used to show that a target belief coheres with other beliefs that have already been accepted. But arguments for individual beliefs have a much diminished role in settling questions of justification.

The goal of this chapter is to introduce a more viable approach to critical thinking, based on the idea of coherence and more closely related to the way people actually reason. But there is

an initial obstacle. The underlying foundationalist intuition seems reasonable: Inference cannot originate justification but can only *transmit* it. There must be a set of basic beliefs for which *Why* questions do not arise.[81] Coherentism is sometimes characterized in terms of what it rejects, i.e., the claim that there are basic beliefs whose justification is independent of other beliefs and which are the ultimate sources of justification for other beliefs. Coherentists seem to be stuck with an absurd claim, the exact converse of the foundationalist intuition: Inference cannot transmit justification from one belief to another, but must *originate* it. According to pure coherentism, *all* beliefs are justified by their inferential relationships to *all* other beliefs. What then is the coherentist solution to the infinite regress of *Why* questions? How do coherentists avoid the pitfalls of skepticism (the prospect that the regress never ends at all), relativism (that it stops at arbitrarily chosen assumptions), or circular reasoning (that it circles back on itself)?

The coherentists' response is a holistic view of justification. They regard the system of interconnected beliefs as a web that is *justified as a whole* by the interconnections within it (Figure 23). These interconnections may be loosely referred to as inferential, but they do not correspond literally to inferential steps in a series of arguments. Rather, such logical, causal, and conceptual relationships enhance the coherence of the belief system as a whole and thereby justify it. Coherentists reject a key implicit assumption, that justification of non-basic beliefs is accomplished primarily by serial *argument*. Circular chains of argument are a problem only if conclusions must be justified by derivation from previously accepted evidence (Day, 1989). Coherentists acknowledge that a chain of arguments for a conclusion might, if pursued long enough, arrive back at the conclusion, just as a chain of dictionary definitions might circle back to the original word. Coherentists find this reassuring rather than troubling, since large circles of argument trace the inferential ropes that bind the entire belief system together. The "conclusions" and the "evidence" of specific inferential relationships are jointly justified by virtue of their membership in a justified system that includes both (Bonjour, 1985, p. 90). The rejection of basic beliefs as a solution to the regress problem leads to a holistic concept of justification.[82]

The holistic view of justification is also the solution to another problem: resolving conflicting arguments. As we showed in the previous section, conflict cannot be resolved by linear argument or step-by-step derivation, since each side (or neither side) might be able to create acceptable arguments against the other. All but the most trivial cases of conflict resolution demand explicit or implicit comparisons of alternative *sets* of beliefs.[83] Because of its inability to handle conflicting evidence, foundationalism is not *sufficient* for justification. Even if there *were* basic beliefs, coherence would have to be called on for at least part of the justification of other beliefs.

Some contemporary foundationalists have adopted hybrid views that acknowledge the role of both basic beliefs and coherence in justification. But the foundationalist aspect of this compromise is weak (Haack, 2000). The traditional rationale for foundationalism is that basic beliefs are needed to save justification from the infinite regress of reasons. But once the

---

[81] This refers to the regress argument for foundationalism: If someone asserts P, a critic may ask for reasons, then reasons for those reasons, and so on, unless beliefs exist which require no justification.

[82] Coherentists need not deny that circularity in argument is a fallacy. *Short* circles – e.g., giving Q as a reason for believing P and P as a reason for believing Q – will be rejected because they fail to reveal inferential relationships that bind P (and Q) to the rest of the belief system.

[83] This is the kind of process studied by psychologists (e.g., Hastie, 1993) who find that jurors reach a verdict by creating and evaluating *stories*.

contribution of coherence to holistic justification is acknowledged (because of its role in conflict resolution), the regress argument crumbles. Basic beliefs are not necessary to stop an infinite regress because justification is not transmitted by argument, and therefore there *is* no regress. Since we had trouble making sense of basic beliefs in any case (because of defeasibility), coherentism is left holding the field. At the very least, the defense of the "foundationalist" element in a hybrid theory cannot depend on the regress argument. A more sophisticated compromise must grant a role for coherence as part of the justification for all beliefs, including so-called "basic" ones.[84]

Although some beliefs may be more firmly entrenched than others, ultimately they are all in the same boat with respect to justification. For any belief (even those based on perception, immediate memory, or logical intuition) there are circumstances in which we might reasonably demand justification for it and eventually reject it because of incoherence with other beliefs. Conversely, virtually any belief (no matter how theoretical) might be accepted without challenge and used as a basis for reasoning about other beliefs. Reasoning is context-dependent: In every context of reasoning, *some* beliefs will be taken for granted – since everything can't be challenged at once – and some claims will be subject to scrutiny. But beliefs that are "foundational" on one occasion may be challenged on another occasion. There is nothing absolute or permanent about their status.[85]

Coherentism is not just a fall-back position, to which we retreat after the failure of foundationalism. Coherentism is motivated (to a greater extent than foundationalism) by observation of the way people in fact reason; as a result it presents a more plausible view of belief change and the growth of knowledge. Belief systems become increasingly coherent through the natural human propensity for pattern recognition. By detecting order in the complex flux of events, humans develop efficient representations or schemas for comprehending, remembering, predicting, and controlling events. Pattern recognition continues to operate at higher levels, finding regularities across schemas that apply in diverse situations, unifying such schemas into richer and more comprehensive structures in which common principles are applied to an increasingly wide diversity of cases. At every level, coherence spurs the generation of more concise descriptions / schemas / theories of ever larger parts of the belief system. At the same time, the growth of structure sets up interdependencies across the knowledge base. These interdependencies make it possible for unexpected information to initiate changes that sometimes (though rarely) ripple widely through the belief system. In other words, coherentism explains how radical conceptual change might take place. Foundationalism, by contrast, supplies no motive to add beliefs except accumulation for its own sake, no incentive to unify or simplify knowledge, and no provision for change except by small increments.

Here is a normative definition of critical thinking from a coherentist point of view.

---

[84] The real motivation for a foundationalist component is the special role that sensory inputs play in justification. We will discuss that issue at the end of this chapter..

[85] Sometimes the coherentist position is expressed as the denial that anything is pertinent to justification except beliefs. As Davidson (1986, p. 310) says, "What distinguishes a coherence theory is simply the claim that nothing can count as a reason for holding a belief except another belief. Its partisans reject as unintelligible the request for a ground or source of justification of another ilk" – such as raw sense experience or a special class of beliefs that directly reports raw sense experience. If sense experience does not have propositional content (i.e., is not the kind of thing that can be true or false), it cannot serve as a reason for a belief. If it has propositional content, then it *is* a belief. And any belief can be overriden by incoherence with other beliefs.

Normative Definition of Critical Thinking #8.: Coherentist

Purpose	*To increase the chance of accepting justified beliefs.*
Constraints	*(1) Systems of beliefs rather than individual beliefs are the units of justification.*
	(2) The justification of a system of beliefs depends on its degree of coherence.
	(2) The coherence of a system of beliefs is based on the inferential relationships among the beliefs within it.
Functions	*Critical thinking is:*
	(1) the identification of alternative candidate systems of beliefs,
	(2) the identification of inferential relationships among the beliefs in the alternative systems,
	(4) identification of coherence-determining characteristics of those inferential relationships,
	(3) acceptance of the system that is most coherent, and
	(4) acceptance or rejection of individual beliefs based on their membership in the accepted system of beliefs.

Clearly, some important questions remain to be answered by coherentists. In particular, do realistic computational limitations allow for the identification of sufficiently coherent overall systems of beliefs? What are the specific characteristics of inferential relationships that serve as criteria of coherence? Can coherentism account for the special status that cognizers give to perceptual beliefs?

Is Coherentism Computationally Feasible?

Directionality is built into the standard definition of argument (Table 2) as a set of statements divided into two subsets: premises that we already accept plus a conclusion that is derived later (and which may then become a premise in a subsequent argument). For coherentists, this appearance of directionality is a by-product of a more fundamental, essentially non-directional evaluative process, which "argues" for and against groups of beliefs, perhaps even our entire belief system. Harman describes this new way of looking at inference as –

> …a way of modifying what we believe by addition and subtraction of beliefs. Our "premises" are all our antecedent beliefs; our "conclusion" is our total resulting view. Our conclusion is not a simple explanatory statement, but a more or less complete explanatory account (1973, p. 159).

On this extreme holistic view, inferential reasoning takes everything we believe as a starting point, and makes changes in beliefs to increase overall coherence.

Not surprisingly, computational tractability is a major potential problem for the holistic view of justification. Foundationalists avoid intractability by focusing on argument in the small, i.e., building up a belief system by many small steps in support of individual beliefs. Foundationalists make this work by imposing three constraints: Modularity of the set of beliefs regarded as relevant in any particular argument, independent evaluation of the premises and

inferences in an argument, and treatment of conflicting evidence in the same way as non-conflicting evidence. Coherentism is a response to the failure of all three constraints in the face of Type C challenges (i.e., conflicting information).

A belief system can be thought of as a complete set of answers to a large set of questions (or equivalently as a complete set of positions on a large set of issues). An exhaustive algorithm would examine every possible belief system (i.e., every possible combination of answers to all the questions), assess the degree of coherence of each system, and select the highest scoring system. The total number of belief systems to be examined is an exponential function of the number of questions. Unfortunately, even for relatively small belief systems, the combinatorics of this approach far exceed human cognitive capabilities (Thagard, 2000, pp. 26-28). In a memorable passage, Cherniak (1986, pp. 93, 143) considers the evaluation of a system containing only 138 questions, each of which has only two answers (*yes* or *no*). Suppose a supercomputer examines one line of a truth table (i.e., one of 2^{138} possible combinations of truth value assignments for 138 sentences) in the time required for a ray of light to traverse the diameter of a proton. The calculations would take longer than the history of the universe to date.

Since an exhaustive search in the space of belief systems is well out of reach, we need to consider other methods for computing coherence, in order to both understand actual human reasoning and to develop meaningful normative criteria. These methods will necessarily be heuristic rather than algorithmic; that is, they will not guarantee discovery of the most coherent system, but may reliably provide close enough approximations. At least three types of strategy can be exploited: mental models or stories, parallel constraint satisfaction networks, and causal structure. All three in combination may bring coherence within practicable reach, and all three involve externalist assumptions.

1. *Mental models or stories.* The first strategy is to consider bite size chunks, i.e., to evaluate the coherence of small subsets of the belief system. For example, Pennington and Hastie show that jurors organize information presented in a criminal trial by means of *stories*. They construct stories corresponding to innocence and guilt respectively, and arrive at a verdict by evaluating the coherence of the competing stories. In constructing these stories, cognizers draw on background knowledge, but the stories do not represent the entire belief system of the juror. Similarly, Johnson-Laird and Byrne claim that reasoners use their understanding of language and their background knowledge to construct *mental models*. Each mental model represents a possible state of affairs with respect to a small set of variables, and is far from an exhaustive description of the entire belief system. According to both Pennington and Hastie, and Johnson-Laird and Byrne, when cognizers evaluate the coherence of stories / mental models, their judgments must draw on background knowledge of relationships among the variables that are explicitly represented in the model / story. Nevertheless, maximizing the local coherence of a model / story in the light of background beliefs is not the same as maximizing the coherence of the belief system as a whole. First, the background beliefs themselves are not subject to revision, and second, it is possible that some relevant background beliefs will not exert appropriate influence on coherence judgments. Thus, there is no guarantee that the most coherent bite-size *story or model* will be part of the most coherent *overall view*.

The limitations of this strategy can be mitigated somewhat by adopting a sequential approach to evaluation. The belief system may be explored one segment at a time by a series of mental models or stories that focus on different (but possibly overlapping) subsets of variables. Each story is adjusted to improve its coherence with respect to the current set of background

beliefs (i.e., beliefs not made explicit in that story). As the cognizer cycles her attention through a series of such models, more background knowledge will be brought into play, and the overall coherence of her belief system should improve. Nevertheless, suboptimal overall solutions are still likely, both because coverage of the belief system will inevitably be incomplete, and because of order effects arising from such a serial strategy.[86]

2. *Parallel constraint satisfaction networks.* Another, complementary method is to utilize computationally feasible mechanisms for approximating maximal coherence over very large sets of beliefs. Prominent among such algorithms is parallel constraint satisfaction in a belief network (e.g., Thagard, 2000; Shastri & Ajjanagadde, 1993; Shastri, 1999a,b). Such algorithms, if they exist, operate non-consciously, in contrast to the explicit consideration of mental models considered above. Parallel processing thus introduces an externalist component into justification. Factors relevant to the justification of a cognizer's beliefs will not be accessible to her awareness. A significant component of reasoning is placed outside consciousness.

Constraints consist of positive or negative links between beliefs, and activation of one belief influences activation of another belief as a function of the sign and weight on the link between them. In Thagard's system, the process starts with a particular set of activation levels across all nodes in the network, and these are adjusted by spreading activation. According to a more psychologically realistic model (Shastri, 1999a,b), activation is initiated by perceptual inputs or by conscious reflection, and then spreads to other nodes from its point(s) of origin; some activation is also contributed by priming due to previous states of the system. The network is likely to settle into a state in which constraints are largely satisfied – that is, beliefs joined by strong positive constraints tend to be either both active or both inactive; and pairs of beliefs joined by strong negative links tend to have one member active and the other member inactive. Such networks settle on a final activation pattern in time proportional to the number of nodes in the network. Computation is thus quite feasible.

A price is paid, however, for computational feasibility. Parallel constraint satisfaction networks may produce suboptimal solutions for two reasons. First, there may be order effects which arise as influence spreads from one part of the system to another; as a result, the final pattern will be influenced by the initial state. Second, in a psychologically plausible realization of such a network, not all knowledge will be accessed with equal effectiveness. There appear to be limits on how far activation can spread, or influence can be exerted, from any particular starting point, and these limitations on spread of activation will exacerbate order effects. Order effects can be mitigated but not eliminated by randomly perturbing activation levels while the network is settling, e.g., as in simulated annealing algorithms (McClelland, Rumelhart, & the PDP Research Group, 1986).

A way to address both order effects and limits on spread of activation is to focus sequentially on parts of the belief system, as in strategy 1 above. This remedy involves (i) serial consideration of different subsets of nodes in the belief system to counter limits on spread of activation, and (ii) serial consideration of different assignments of truth and falsity to under-constrained nodes to counter order effects. By shifting attention among specific subsets of beliefs in a large network, cognizers will increase the span of operation of the automated constraint

---

[86] Lehrer's (2000) approach to coherence is a sequential strategy of this kind, but instead of stories or models, one belief at a time is singled out for evaluation. Lehrer, however, does not treat this strategy as an approximation to a more ideal one, and does not address its rationale in terms of limited computational capacity.

satisfaction mechanism and the amount of background knowledge that is brought to bear on the problem (persistent priming is one vehicle of integration across attentional cycles). Consideration of alternative truth value assignments to unconstrained nodes may reveal system states that are relatively inaccessible by automated spread of activation given the initial state of the system, but which are nonetheless highly coherent. If conclusions are unstable across variations in attentional focus and activation values, further exploration of alternatives and collection of information may be warranted until the solution stabilizes.

A computationally feasible implementation of coherence is likely to operate at two levels, one deliberate and conscious and the other automatic and non-conscious. Strategy 1, conscious consideration of mental models, and strategy 2, automated constraint satisfaction by activation in a belief network, are both necessary. From the point of view of critical thinking theory, this makes sense. The acknowledgement of strategy 2 breaks with the internalist insistence that all aspects of reasoning be accessible to awareness. On the other hand, Strategy 1 allows us to identify an *optional* reflective mode of thinking that is synonymous with reasoning as such. The process of generating mental models by explicitly introducing new issues and considering alternative possibilities is *critical thinking*. Conscious reflection on mental models or stories can improve the degree of coherence attainable by automatic spreading activation, by mitigating order effects and overcoming limitations on the spread of activation. Conversely, conscious reflection is possible only in the context of a more comprehensive non-conscious system of background knowledge (Cohen, Thompson, Adelman, Bresnick, Shastri, & Riedel, 2000c; Forster, 2001). The picture we end up with is that of a shifting spotlight of attention (i.e., parts of a story or mental model), surrounded by a penumbra of activated background knowledge (the activated portion of long-term memory). Both are contained within a larger region of relatively inactive knowledge to which attention might subsequently be shifted by critical thinking.[87]

3. *Causal structure*. The resolution of conflicts among different sets of evidence or lines of reasoning can involve revision of assumptions anywhere in the system of beliefs, and is not limited to a well-defined segment that can be thought of as "under" the beliefs in question. In the transition from foundationalism to coherentism, *holism* therefore replaces modularity as a basic principle of reasoning (Figure 23; Quine, 1953). Yet strategies 1 and 2 together reintroduce a form of sequential reasoning in order to overcome capacity limitations. As in foundationalism, reasoning is broken down into steps, each of which considers only a subset of the belief system (a conscious mental model surrounded by a region of activated background knowledge). But sequential processing in a coherentist framework is not "linear" in the same way as sequential reasoning in a foundationalist framework. There is no fixed constraint on the ordering of steps, viz., from basic to derived beliefs. Rather, cognizers may adopt a variety of different strategies for shifting attention among different subsets of the total system, improving coherence as they go along, until the results stabilize.

Is there any principled way to carve out "natural" subsets of beliefs for separate, sequential consideration? If the answer were *no*, if for example every part of the belief system

---

[87] For more detail on the implementation of such a system, see Cohen, Thompson, Adelman, Bresnick, Shastri, & Riedel, 2000). In this system, each belief node is associated with "prior probabilities" that summarize the historical activity of nodes upstream from that node in the network. Whenever the node is on the edge of a sphere of activated knowledge, its prior probability influences its activation level. These prior probabilities allow inactive knowledge to have an influence based on historical activity patterns which are not adapted to the current context.

were densely connected to every other part, the sequential process described above would be unlikely to produce a good approximation to optimal overall coherence, and the attempt to formulate sensible strategies for shifting attention would likely fail. In fact, however, modularity survives the fall of foundationalism, although in a different form. The relevant sort of modularity results from stable *causal* relations among *events* rather than from supposedly fixed *evidence-conclusion* relations among *beliefs*. The key principle is that consideration of shared causes simplifies the representation of interdependencies among events. As a result of shared causes, beliefs about causal relations among events can generally be divided into natural, modular subsets.

Figure 24 shows how shared causal relationships can induce modularity. The top of Figure 24 shows a network of event representations. Each event is a binary variable with two values; the event either occurs or does not occur. The occurrence of each event is correlated positively or negatively with the occurrence of each of the others. An exhaustive algorithm would have to consider 2^{12} (= 4096) possible patterns of occurrences and non-occurrences to determine which combination was most coherent. A constraint network with 12 nodes obviates the need for serial consideration of all possibilities. A selection is made automatically as the network settles into a state that (nearly enough) satisfies constraints. Such a network can approximate the ideal solution with only one link parameter for each of the 66 pairs of events (Thagard, 2000).[88] This approximation requires that activation reach all nodes in the network (which Thagard appears to assume is the case). If conscious attention and spreading activation reach only part of the network, the solution may deviate significantly from optimality, for the two reasons mentioned previously: Revisions of belief in the inactive portion of the network will not be possible either through conscious consideration or automatic constraint satisfaction, even though such revisions might be part of a more coherent overall solution; and some constraints will not influence the solution, even though they reflect important background knowledge of relationships among events.

[88] A more complete Bayesian model, however, involves asymmetric links, such as those between cause and effect, and would require 66 x 2 = 132 link parameters.

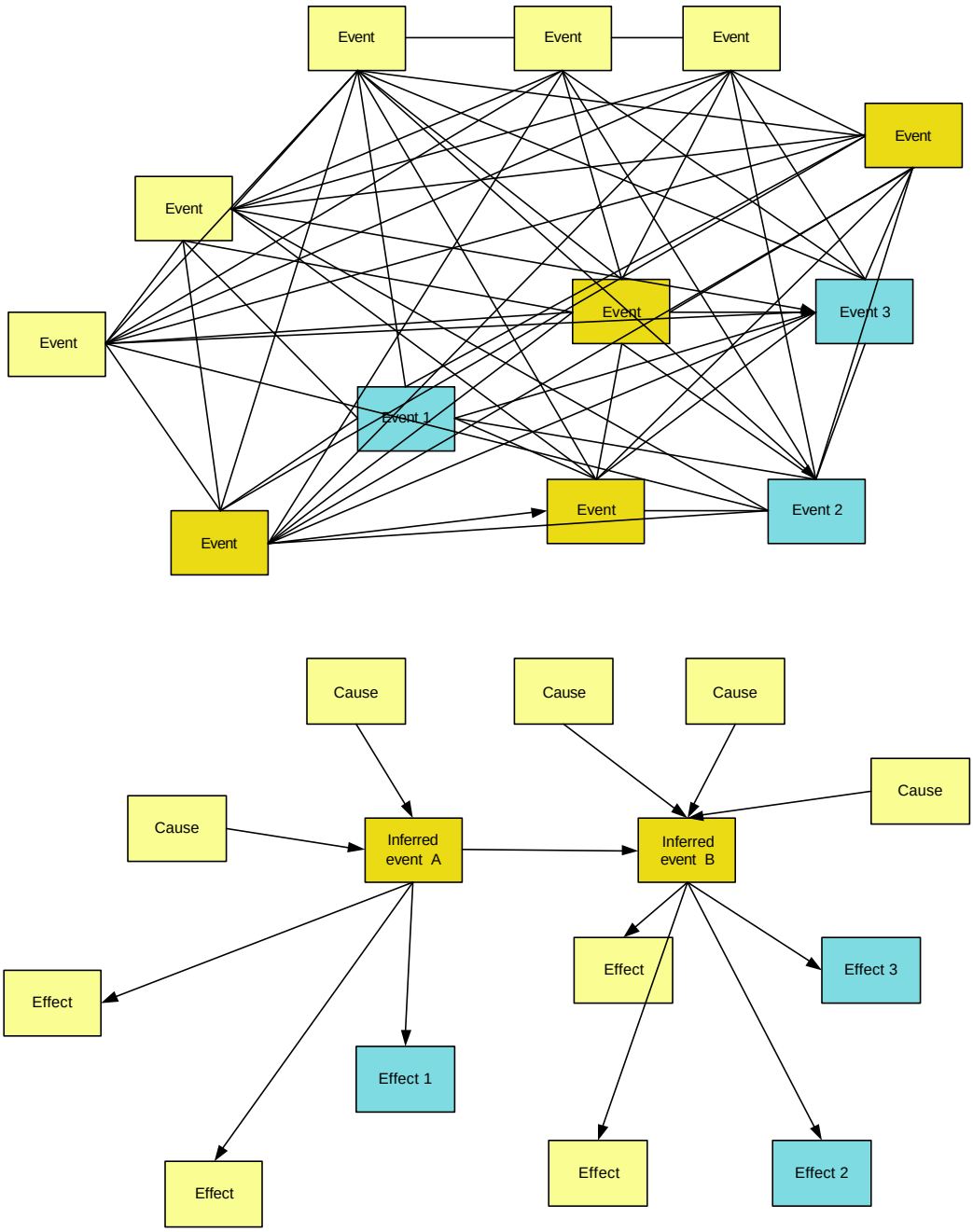


Figure 24. At top, constraints among events are represented by associative links. At bottom, the same events and constraints are represented more economically within a causal structure. Directed arrows represent asymmetric cause-effect relations. Blue nodes (1, 2, and 3) are consciously attended. Gold and blue nodes are activated.

Suppose, for example, that only beliefs 1, 2, and 3 in the top diagram of Figure 24 are under conscious consideration, and that four additional nodes are activated but not under conscious consideration – for a total of seven active nodes. Judgments about the coherence of mental models ideally would draw on the entire belief network. But if only seven out of the 12 nodes actively influence the solution, only 21 out of 66 link constraints will play a role in

selecting the most coherent model (or 42 out of 132 link constraints in a Bayesian model). Moreover, only 2^7 (= 128) mental models are available for selection, which falls far short of the actual total, 4096. Finally, as previously noted, a strategy that involves shifting attention in the belief network will improve the accuracy of coherence judgments, but the solution will still be subject to order effects.

The bottom of Figure 24 shows the same 12 events organized in terms of causal relationships. In this particular causal theory, there are two inferred events, A and B, one of which has a causal effect on the other (for example, the intent to attack at a particular place causes the intent to locate artillery within range of that place). All other events are either causes or effects of A or B. Now, a crucial property of causal relationships can come into play. Suppose event x and event z are correlated, but have no causal connections except those that pass through event A. That is, either x → A → z, x ← A ← z, or x ← A → z.[89] Then if event A is known, it is not necessary to know the status of event x in order to predict the status of z, or to know the status of z in order to predict the status of x. In other words, event A "blocks" the interdependency between events x and z (Sober, 1994). Because of this property, paradoxically, adding one or more shared causes dramatically simplifies the overall associative structure. The two inferred nodes A and B in the bottom part of Figure 24 insulate each effect from the influence of the other effects, and insulate each effect from all the higher level causes. As a result, if we have an appropriate degree of belief in the common causes A and B, no further information about any of the other effects or about the higher level causes is necessary. The result of blocking in the bottom of Figure 24 is that only 26 link parameters (in a Bayesian model) are sufficient to explain all the regularities in the original data. More importantly, if events 1, 2, and 3 in Figure 24 are under conscious consideration, activation needs to spread only to two additional nodes, A and B, in order to capture the influence of *all* the other events. Conscious judgments regarding events 1, 2, and 3 are therefore far more likely to adequately reflect the relevant background knowledge. Moreover, introducing the two inferred causes permits an efficient sequential strategy for finding the most likely candidates for belief revision in the network. If one of the two inferred nodes, say A, is selected for revision, attention may shift to A, leading to the activation of causes and effects of A that were previously inactive. These may now be considered as candidates for revision, and so on.[90]

Causal structure mitigates the limitations of mental models and parallel constraint satisfaction as methods for achieving coherence. By contrast with foundationalism, causally based modularity is not postulated arbitrarily to fit a preconceived conception of linear reasoning. It is justified in terms of *actual* external relationships among events. Unlike foundationalist modularity, however, causal modularity is not perfect. It works only if there are no additional, unrepresented shared causes.[91] A and B account for all the correlations among the effects of A and B only if there are no competing explanations of the effects. Such competing

---

[89] This excludes one possible case: x → A ← z, where x and z are either competing explanations of A or part of the same causal process that produces A. In this situation, x and z are independent (barring some further shared cause), but *become correlated* given knowledge of A (Pearl, 1989).

[90] A more complete description of a system that implements these ideas can be found in Cohen, Thompson, Adelman, Bresnick, Shastri, and Riedel (2000).

[91] For a given node, say A, the effects of the rest of the network are blocked given knowledge of (i) A's causes, (ii) A's effects, and (iii) alternative causes of A's effects (Glymour, 1980). Nodes in (i), (ii), and (iii) are called the *Markov blanket* of A.

explanations are *defeaters* for the explanation of the effects in terms of A and B. There is ordinarily no way to be sure that all alternative accounts, i.e., defeaters, have been eliminated. Thus, modularity must ultimately be regarded as a hypothesis subject to challenges of type C (*Even if A, and A causes B, it is still possible that B will not occur, if some alternative causal influence interferes.*)

Alternative explanations might arise from anywhere in the web of beliefs. Nevertheless, critical thinking with the aim of finding defeaters and exploring likely alternative causes is restricted, at least at first, to a smaller range of beliefs, usually within a particular domain. Beliefs in different domains of knowledge usually have little or no impact on one another. The only links among them might be, for example, high level analogies or a shared methodology, (Quine mentions the use of logic, probability, and mathematics as linking different scientific domains; 1990). These barriers between domains reflect another, coarser form of modularity that is added to the finer grained modularity induced by causal structure itself. Modularity of this kind too, however, is not perfect. Many problems are multi-disciplinary (e.g., attack planning involves knowledge of weather, weapons performance, and human motivation.) It is not all that uncommon for different domains to offer competing explanations of the same event (e.g., Was the slow movement of the enemy convoy due to vehicle capability combined with road conditions, or the need to synchronize with another force?) Where a linkage is not already evident, there is always the possibility of expanding the inquiry to discover unsuspected fundamental causes, principles, or analogies across domains.. In the final analysis, modularity is not absolute, but it does facilitate strategies for efficient search that make a holistic account of reasoning feasible.

Arguments Large and Small

What makes one set of beliefs more coherent than another? Although coherentism has been accused of vagueness on this question (cf., Thagard, 2000, pp. 69-70), criteria of coherence have in fact been spelled out in varying degrees of detail and precision by Lehrer, Bonjour, Lycan, Harman, Haack, Thagard, and Quine. Table 10 summarizes some of the criteria that have been proposed. We have divided them into five categories: logical consistency, mutual support, generality, simplicity, testability, and conservatism. An understanding of these criteria will shed light on a version of critical thinking that stresses the key role of causal explanation.

Table 10. Criteria that have been proposed for evaluating sets of beliefs.

Harman (1986, pp. 55-75)	**Bonjour (1985, pp. 95-99)**	**Thagard (2000, pp. 43 - 63)**	**Lycan (1988, p. 130)**	**Quine & Ullian (1970, pp. 42-53)**
Logical consistency				
	Logically consistent	Fewer contradictory hypotheses		
Mutual support				
	Fewer competing or negatively associated hypotheses	Satisfaction of positive and negative constraints		
Best total explanatory account	Number and strength of inferential connections between component beliefs	Explanatory, analogical, deductive, conceptual, or visual connections		
Comprehensiveness				
	Fewer unexplained anomalies in the belief content of the system		Fewer messy unanswered questions	
	Not divided into relatively unconnected subsystems of beliefs		Explains more	Generality of explanation
Simplicity				
Minimizes clutter, or uninteresting beliefs		Fewer hypotheses required in the explanation	Simplicity	Simplicity
Testability				
			More readily testable	Refutable
Conservatism				
Minimizes addition or subtraction of beliefs			Squares better with what you already have reason to believe	Requires rejection of fewer accepted beliefs

Logical Consistency

Logical consistency is often proposed as a criterion, and sometimes even as a synonym, for coherence. A common objection to this version of coherentism is that there can be more than one logically consistent set of beliefs. This is true. As we have already seen, consistency does not uniquely determine the best set of beliefs even if supplemented by the constraint that reasoning start from acceptable premises. Bt the identification of coherence with consistency (logical, probabilistic, or any other kind) could not be more incorrect. On the contrary, the insufficiency of logical or other kinds of consistency is one of the motivations for rejecting argument-centered views in favor of coherentism. Sound arguments from accepted premises may be constructed for both sides of a conflict (or for neither side, if the threshold of acceptability is set too high). Different chains of argument, each based on plausible premises, can lead to equally consistent but logically contradictory conclusions. To resolve the conflict, bodies of belief must be evaluated and compared as a whole rather than assembled step by step through argument. And evaluations of bodies of beliefs must be based on more than mere consistency.

Mutual Support

What then is it about logical and probabilistic relations that contributes to the coherence of a belief system? It is best to start with the core idea of coherentism, that beliefs are justified if they are interconnected in such a way that they *mutually support* one another. This core idea can take several different, related forms. For example, Haack (1993) uses the analogy of a crossword puzzle, in which each word that is filled out provides new constraints that help us fill out other words correctly, even though none of the words is more "basic" than any other. To the extent that a system is coherent, each accepted belief or set of beliefs tends to increase the likelihood of other beliefs in the system. Bonjour (1985, p. 98) recommends as one criterion of coherence, the degree to which there are positive inferential connections among accepted beliefs. Note that this goes well beyond probabilistic consistency in the sense discussed above. A system might be probabilistically consistent even though it consists of isolated beliefs that have nothing whatsoever to say about one another (i.e., whose probabilities are independent).

An attempt to give a more general, as well as more precise, account of mutual support among beliefs is Thagard's (2000, p. 17; 1992) notion of *constraint satisfaction* in a network. The elements of the network, according to Thagard, include propositions, concepts, actions, and goals. Each pair of elements is joined by a symmetrical positive constraint (e.g., explanation, deduction, positive association, etc.) or by a symmetrical negative constraint (e.g., logical inconsistency, competing explanations, negative association, etc.). A positive constraint between two elements is satisfied when both of the elements are accepted or both are rejected. A negative constraint between two elements is satisfied when one element is accepted and the other is rejected. Inferential reasoning tries to accept and reject elements so as to satisfy the greatest possible number of constraints, weighted according to their degree of positive or negative importance.[92]

---

[92] Thagard's criterion implies not only that constraints should be satisfied, but that they should exist. A system in which no constraints exist will score low on coherence because all the weights will be zero. If acceptance and rejection are a matter of degree, a measure of coherence is given by $\Sigma_i\Sigma_j w_{ij}a_ia_j$, where $0 \leq w_{ij} \leq 1$ is the weight on the link between units i and j, and $0 \leq a_i \leq 1$ is the activation of unit i (Thagard, 2000, p. 32, 38-9). Sensitivity to the number of elements n in a belief network can be eliminated by dividing by n.

The criterion of constraint satisfaction, even in Thagard's detailed formulation, leaves something to be desired: namely, a specification of the *kinds* of connections or constraints that count toward coherence and why they count. Thagard's criterion can be applied only after a set of constraints among elements has been hand-crafted. Thagard provides rules for establishing positive and negative connections of several different types, viz., explanatory, deductive, analogical, perceptual, and conceptual. But it is not clear what the rationale for the rules is – or, more precisely, why a system of beliefs becomes justified through satisfying these particular types of constraints. In the absence of such a rationale, a constraint satisfaction measure of coherence is arbitrary as to what it might include.[93]

In response to this problem, some coherentists (e.g., Harman, 1973; Lycan, 1988; Quine & Ullian, 1970; Bonjour, 1985) have adopted an *explanationist* position. The explanatory role of beliefs suggests that whatever features characterize good explanations might also serve as criteria of coherence and therefore of justification. A second reason for stressing the priority of explanation is that explanation is closely linked to coherence construed as mutual support. Explanation involves finding a unitary account of a disparate set of phenomena, and such a unitary account shows how the phenomena mutually constrain one another. A third reason for stressing the explanatory role of beliefs is the fact that causal explanation induces modularity in a belief system. Such modularity is what makes possible the sequential consideration of smaller subsets of beliefs, i.e., mental models or stories, and sequential consideration of belief subsets is what makes the evaluation of coherence computationally feasible. If this point of view is correct, coherence is increased by processes that operate at several different rates: (i) the relatively slow learning processes that we use to extract causal relationships from our experience (Pearl, 2000; Glymour, 1980), (ii) the somewhat faster, but still deliberative processes of critical thinking about such relationships, and (iii) the rapid processes of constraint satisfaction that operate during automatic inferencing about such relationships.

What virtues must a good explanation have, and can they also serve as criteria of coherence? A number of evaluative criteria have been suggested, as shown in Table 10. Good explanations are (i) comprehensive and (ii) simple (Figure 25), but they should also be (iii) testable and (iv) as compatible as possible with existing beliefs.[94]

---

[93] This is reminiscent of the ambiguity in informal logic regarding the "acceptability" of premises and "sufficiency" of inferences. Accounts of these usually boil down to lists that lack systematic rationale.

[94] How do explanationists deal with constraints among beliefs that do not appear to involve causal explanation, such as analogical, semantic, logical, or mathematical relations? Harman (1973) and Lycan (1988, p. 178) argue that these types of constraints are also explanatory in a broader sense; they increase the intelligibility of other phenomena and thus can be said to explain them. The explanationists' proposal, in effect, is not to exclude non-explanatory kinds of reasoning from the scope of coherence theory, but quite the contrary, to generalize the application of criteria for causal explanation to other purposeful thought that might at first appear non-explanatory. As Lycan declares (p. 125): "We are always and everywhere stuck in the business of making comparisons of plausibility, and such comparisons are made only by weighing explanatory virtues" (e.g., simplicity and comprehensiveness).

Thagard may be a crypto-explanationist. He proposes different rules for each of six different types of coherence, viz., explanatory, deductive, analogical, perceptual, conceptual, and deliberative. This is troubling, since it suggests that there are six different theories rather than a single "coherent" concept of coherence. However, Thagard (pp. 62-63) suggests that there are parallels among the different sets of rules. In particular, there are rules corresponding to comprehensiveness and competition among alternative views for all types of coherence. The role of simplicity is made explicit only in rules for explanatory and deductive coherence, although Thagard (p. 64).remarks that simplicity also is probably more general. Thagard does not ask why these parallels exist, but the shared features of his rules seem to correspond to explanatory virtues. This commonality can be taken as further

What Questions Does a Story Answer?

Comprehensiveness

Comprehensiveness refers to explanatory power, which we will take to include (i) the range and variety of the phenomena that are explained, (ii) the level of detail or specificity of the features that are accounted for, and (iii) the accuracy of the fit between the phenomena and the explanation.[95] For example, MAJ Nord's hypothesis that the enemy intends to attack in the north in the near future can explain or predict a variety of other beliefs, which pertain to gross features of the location and timing of enemy troop movements, composition, and concentrations, enemy reconnaissance activities, radio traffic, engineering operations, and placement of artillery (e.g., beliefs #1 - #5 at the bottom of Figure 25). If one of these activities is anomalous (e.g., the placement of artillery conflicts with the explanation), then comprehensiveness is reduced. It is also reduced, though to a lesser extent, if there are other enemy activities about which the hypothesis of intent to attack makes no predictions at all. A more detailed hypothesis about enemy actions (e.g., based on more specific knowledge of the commander's preferred tactics) might result in a more comprehensive account, which covers anomalies or apparently irrelevant events.

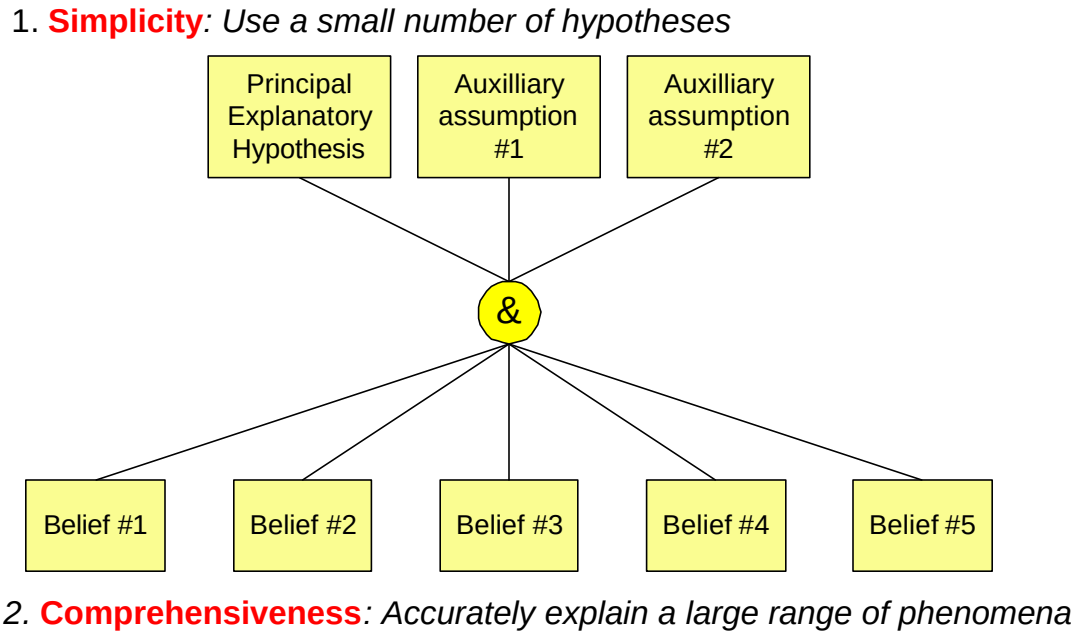


Figure 25. Simplicity and comprehensiveness as criteria of the coherence of an explanation.

According to both foundationalism and coherentism, the phenomena that are explained support the truth of the explanation. But for coherentism, support runs in the other direction as well: The existence of an explanation provides some support for beliefs about the phenomena.

---

support for the explanationist position. See footnote 96.

[95] As long as we are in a pure coherentist framework, when we speak of phenomena or data, we mean observational *beliefs*.

We may be less confident about the occurrence of activities which conflict with our explanations of other events, or for which we have no explanation at all. One way to improve comprehensiveness in a coherentist framework is to discount observations that are not well explained. For example, if the location of enemy artillery does not fit well with the best supported hypothesis about enemy intent, we may question our information about artillery location. Typically, however, we demand an *explanation* for the decision to disregard an apparent datum. When no enemy artillery is seen in the area where an enemy attack otherwise seems imminent, it would not be unreasonable to look further into the possibility of concealment, poor coverage of the area by scouts, dishonest informants, use of other means (such as air power) to accomplish the function of artillery, or longer range artillery placed elsewhere. In the absence of such an explanation for discounting the evidence, comprehensiveness declines after all. But notice that explaining away comes at a price: In the absence of independent evidence for the explanation, the *simplicity* of the overall account is sacrificed.

Simplicity

Comprehensiveness is empty without simplicity. For example, if an ad hoc explanation is offered for apparent anomalies or irrelevant events, then comprehensiveness is restored (i.e., every event has some explanation), but there is a corresponding reduction in simplicity. If there were no bounds on complexity, we could generate an endless variety of theories to predict any and all observations or apparent observations. For example, one such "theory" would postulate an independent cause for each event or event feature to be explained. Simplicity (or parsimony) decreases with the number of different explanatory accounts that must be utilized, and with the number of auxiliary hypotheses (at the top of Figure 25) that must be introduced to make a particular explanation work. The more a schema must be elaborated in order to account for phenomena, the less well it actually explains them.[96]

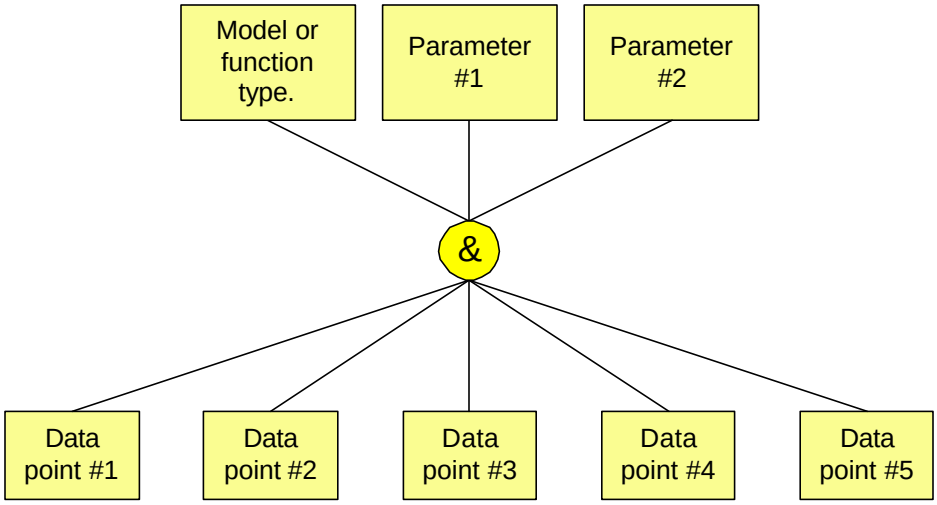
The importance of simplicity is often illustrated in terms of fitting curves to data (e.g., Kuhn, 1996). Figure 26, which parallels Figure 25, depicts curve fitting as a type of explanation. Thus, the principal hypothesis at the top of Figure 26 postulates a particular type of mathematical function or model, and the auxiliary hypotheses state the values of its adjustable parameters. The beliefs at the bottom state the values of the observed data points that are to be accounted for by the hypotheses. Any finite set of data points can be fit exactly by an infinite number of different curves; the data can be "explained" by an infinite number of different mathematical functions as long as they are associated with a sufficient number of adjustable parameters. There is no telling which mathematical function is best in terms of comprehensiveness since each one can fit the

---

[96] Thagard (2000)'s rules for evaluating explanatory coherence include efforts to capture both comprehensiveness and simplicity (as noted in footnote 94). In particular, comprehensiveness corresponds roughly to Thagard's rule that data and hypotheses in an explanation should be positively linked to one another (Thagard, p. 43). Thus, coherence is reduced by anomalies which are not causally linked to other beliefs. Thagard tries to capture simplicity by proposing that the weights on links among hypotheses are decreased as the number of hypotheses in an explanation increases. Finally, by joining competing explanations by negative constraints, Thagard captures the idea that alternative explanations compete with one another, and forces a choice among different possible accounts. The real content of the coherence theory lies in these details, but Thagard does not offer a systematic rationale for them. Moreover, the rules do not quite do the job. For example, coherence may be maximized by a system in which there are no explanatory hypotheses at all (maximum simplicity!), but in which all observations are positively or negatively correlated directly with one another, as in the top of Figure 24. Thagard's criterion does not reward the role that causal explanation plays in simplifying the network as a whole, as in the bottom part of Figure 24.

observed data equally well (i.e., perfectly). The only way we can choose among the various models is by simplicity.

1. **Simplicity**: *Use a model with a small number of adjustable parameters*



2. **Comprehensiveness**: *Fit as much data as possible as accurately as possible.*

Figure 26. Simplicity trades off with comprehensiveness in curve fitting.

The preceding line of reasoning suggests that simplicity is inescapable but also, paradoxically, not very important. If all the curves or theories fit the data points equally well, variations in simplicity are bereft of real empirical significance. Simplicity appears to be nothing more than an arbitrary way of choosing among equally comprehensive / accurate theories or curves, and this might be done in different ways. Different simplicity metrics may be based on the number of hypotheses, objects, object types, concepts, or variables in the account (cf., Swinburne, 1997; McAllister, 1996). Different people will put different weight on different dimensions and thus judge simplicity differently. But because there are no empirical consequences, there is no real right or wrong in these judgments. It is merely a matter of taste.

It seems clear, however, that there is more to simplicity than convenience and aesthetic preference. Forster and Sober (1994; Forster, 1999a, 2001) point out that the preceding rationale drastically understates the role of simplicity in curve selection (and, by analogy, in theory evaluation). This rationale cannot explain why, under common conditions, simplicity leads us to prefer mathematical functions that do *not* fit the observed data perfectly over those that do (Figure 27).

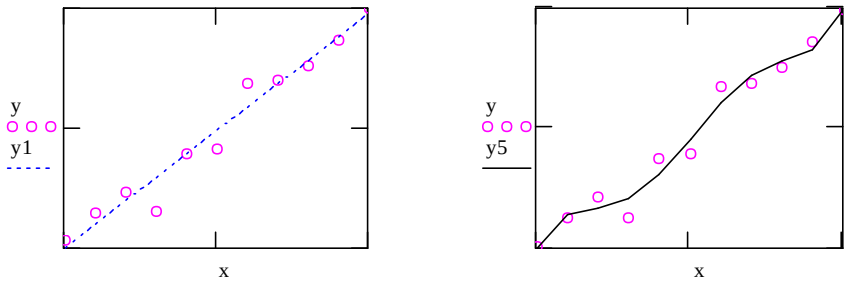


Figure 27. The same data points (y) are fit by a simple curve on the left (y1, a linear function with two parameters) and by a complex curve on the right (y5, a fifth-order polynomial with six parameters). Although y5 fits the observed data better, y1 will usually be preferred.

As illustrated in Figure 27, we do not typically prefer the curve the fits the data best. Perfect fit can of course be achieved by increasing the number of parameters. But the accuracy - simplicity tradeoff kicks in as the number of parameters grows larger well before perfect fit is achieved. At some point, adding parameters to get a more precise fit will not be worth the price in increased complexity of the curve (or theory). We will prefer a simpler function that does not fit the data quite as well. Simplicity, it appears, is not a secondary criterion that is applied only when there is a tie in comprehensiveness. Simplicity is a primary criterion that trades off against comprehensiveness in the choice of the best overall explanation (Quine & Ullian, 1970).

One reason for the importance of simplicity is pragmatic utility. We saw earlier that postulation of causes simplifies the explanatory account necessary for prediction since causes summarize the influence of more distant parts of the belief network. In particular, once the cause is known, an effect can be predicted without knowledge of other effects or of causes further upstream. The same consideration argues for simplicity *within* the description of the causal mechanism itself. Regardless of how simplicity is interpreted (e.g., number of hypotheses, objects, or concepts), simplicity is correlated with *conciseness* of the overall theory or schema.[97] This idea provides a common metric and a uniform account for all types of simplicity (see footnote 94). The aim is to achieve the shortest possible overall description – whether this is achieved by providing a causal explanation or by economizing on objects, types of objects, concepts, parameters, and/or hypotheses within that explanation. Coherence is desirable in part at least because the combination of comprehensiveness and simplicity is cognitively *efficient*. A good explanation compresses large amounts of information into a concise description. It summarizes many beliefs in terms of a small number of economical schemas or patterns. A more concise theory generally imposes less burden on memory, attention, and communication.[98]

---

[97] Gell-Mann (1994) defines the *crude complexity* of any system as "the length of the shortest message that will describe [it], at a given level of coarse graining [i.e., detail]…." A simple theory of a system minimizes the length of the message required to describe it by capturing redundancies or regularities.

[98] Sometimes it seems that the most concise theory requires more effort to apply. For example, the most concise theory will probably not be the most usable if, when it is first used, it is involves less familiar concepts and methods. As Gell-Mann (1994) remarks, however, it is necessary to include within the measure of length any explanatory notation that is required to enable people to understand and apply the new theory. Familiar theories will not be so burdened – and an unfamiliar theory will have to overcome this disadvantage by gains in conciseness (or other explanatory virtues) elsewhere.

There is reason to expect a general correlation between familiarity and conciseness. Indeed, the effect of familiarity on cognitive efficiency may be mediated, in part at least, by its effect on conciseness of representation.. Language evolves to maximize average conciseness of utterances (cf., Zipf's law), and to do this it assigns the

An especially important dimension of simplicity, from the pragmatic point of view, is the number of auxiliary hypotheses (Figure 25) or parameters (Figure 26). Minimizing the number of parameters or auxiliary hypothesizes has effects that go beyond merely reducing the length of the description. Each auxiliary hypothesis or parameter is a *question* that must be answered by consulting the data in order to make the explanation complete. Thus, the more free parameters or unconfirmed auxiliary hypotheses an account has, the more information is needed to apply the theory (Sober, 1975), and the more work must be done before the theory can be *used* in any particular situation. As shown in Figure 28, in a comparison between two theories with the same empirical coverage, the simpler theory uses up fewer of the observations just to specify the auxiliary hypotheses or parameters, and therefore yields more predictions for a given amount of effort (Quine, 1960, p. 20; Lycan, 1988, p. 141).[99] More generally, we prefer theories that require less time or effort over theories that for whatever the reason require more – whether due to the size of the theory, its unfamiliarity, or its requirements for information collection, calculation, and reasoning (Harman, 1999a, pp. 83-85). According to this pragmatic view, improving coherence is a matter of maximizing results (e.g., the prediction of future events) while minimizing the effort required to achieve them.[100]

---

shortest linguistic expressions to the most frequently used concepts. Thus, familiar concepts will tend to be more efficiently representable, i.e., more concise, while novel concepts will at first demand lengthier descriptions. As a theory is used over time, it will become more concise, hence, "simpler." Simplification over time will also be produced by two closely related psychological phenomena: (i) chunking multiple concepts into a single unified representation (Newell & Rosenbloom, 1981) and (ii) compilation of multi-stage reasoning processes into automatic routines (Anderson, 1982). As we shall discuss later, the role of familiarity implies that simplicity / conciseness is *path-dependent*.

[99] Figure 28 ignores error in the observations. In the presence of such error, however, the same conclusion holds (e.g., Figure 27). To achieve a given level of reliability in the estimate of each parameter value, the more free parameters there are, the more data points must be sampled.

[100] As we saw in note 98, conciseness in notation counts toward simplicity, in part at least, because it translates into reduced computational effort. Thus, it makes sense to measure simplicity directly in terms of computational effort rather than conciseness (which is only one contributor to computational effort). Such a measure will aggregate difficulty of storage, attention, and communication (which are all related to length) with difficulty of comprehension, recall, information collection, calculation, and reasoning (which are related to familiarity). Along the same lines, Harman (1999a, p. 86) suggests a combination of size and number of computational steps as a measure of simplicity. These measures do not presuppose that there is a single reservoir of "mental effort" or processing capacity. They do presuppose that expenditures of different kinds of cognitive resources can be converted into a common currency (i.e., a weighted combination) that reflects the relative value of different types of resources. It is clear that such relative values will vary with tasks and contexts.

Note that although familiarity may influence simplicity, it is not necessarily the deciding factor. A theory with fewer auxiliary hypotheses and parameters may be simpler *on the whole* despite introducing unfamiliar concepts. Application of such a theory will involve asking and answering fewer questions, even though the effort involved for each question may be somewhat greater. And, once adopted, the new concepts and methods will eventually become familiar, thus more concise and more efficient.

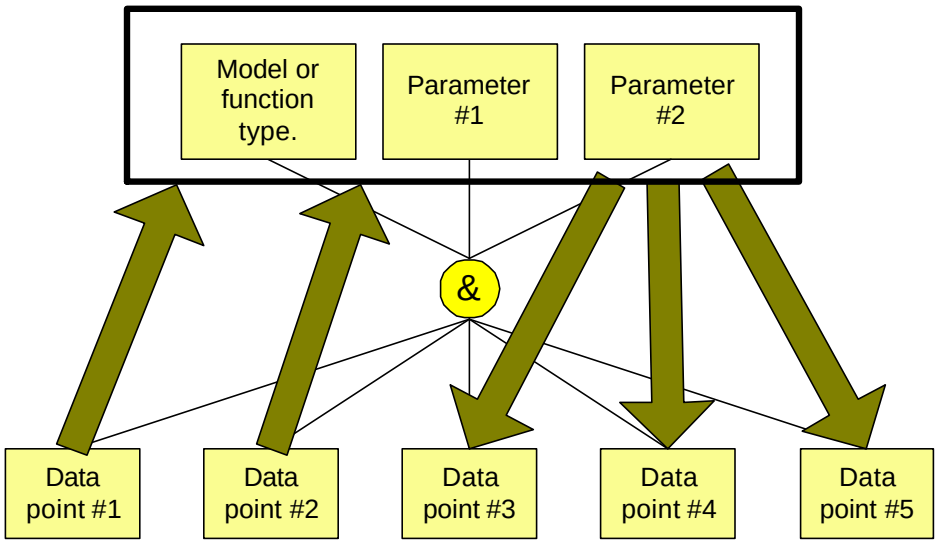


Figure 28. A model with two adjustable parameters uses up two data points to estimate the parameters. Thus, it predicts only three out the five observations.

These pragmatic advantages of simplicity – in terms of both length and information requirements – are extremely important, but they give us no reason to suppose that simpler theories are more likely to be *true*, or are a better approximation to the truth, than complex theories. So, the question arises, whether simplicity has a language-independent, *empirical* significance that goes beyond aesthetic taste or pragmatic utility. In support of this, we note that the case for simplicity thus far leaves out an important fact: Two models that fit beliefs about current data equally well but differ in simplicity may also differ in their predictions regarding *future data*. In such cases, simplicity might in principle serve as an indicator of successful predictive performance. A case could be made that scientists treat it as such an indicator, since they (along with everyone else) prefer to base predictions on simpler theories even when the more complex theory is quite tractable (Swinburne, 1997). But it is difficult to see how simplicity could endow a prediction with extra predictive reliability over and above what is already implied by the fit of the model to the observed data.[101]

Forster and Sober (1994; Forster, 1999a, 2001) provide a rationale for an empirical role of simplicity in curve fitting based on a statistical framework developed by Akaike. When there is random error in the observations, simpler mathematical functions are likely to be closer to the truth – even though they fit beliefs about *observed* data less well than more complex models. The reason is that the complex model (e.g., the right-hand curve in Figure 27) is likely to *overfit* the observed data: As the number of parameters increases and/or as the random error in the observations increases, the more likely it is that adjustable parameter values reflect chance features of the data rather than an underlying regularity. A formula that balances fit to current data against simplicity (e.g., the left hand curve in Figure 27) will do a better job predicting *future* data generated by the same process, because it is likely to be closer to the true underlying process. This is illustrated in Figure 29. Simplicity trades off against fit to data by taking into

[101] Some "solutions" to this problem simply beg the question, e.g., by asserting that simplicity is an indicator of truth because "nature is simple" or because simplicity is an a priori constraint on cognitive representations.

account the *error* in the observed data and avoiding adjustments that are likely merely to fit accidental features of that data. The empirical advantage of simplicity is more accurate prediction.[102]

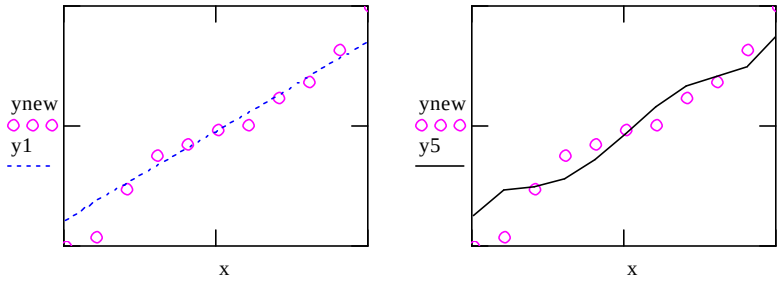


Figure 29. When the two curves in Figure 27 are used to predict new data generated by the same process, the results are reversed: The simpler, linear curve (y1, on the left) fits the *new* data better than the complex higher-order polynomial (y5, on the right).

The predictive utility of simplicity is probably not limited to curve fitting. The same tradeoff between comprehensiveness versus simplicity arises with schemas (i.e., informal theories) that people use in everyday reasoning. For example, MAJ Nord's hypothesis about enemy attack is relatively simple, but it nevertheless involves two parameters, time and place of attack, each of which must be estimated from current observations with an appropriate degree of resolution before any predictions are possible. Moreover, to accommodate apparent conflicts or anomalies, the theory may be made more complex (comparable to adopting a higher-order polynomial to get a better fit to the data). For example, when the location of enemy artillery does not fit her hypothesis about the location of attack, MAJ Nord complicates the theory by introducing another "adjustable parameter," the range of enemy artillery. By assuming that enemy artillery range is greater than expected (without any independent evidence that this is the case, such as prior intelligence reports), she is able to obtain a fit between the theory and the data. A few such complications may be tolerable, especially if the theory successfully explains a large range of beliefs. (In the curve fitting case, after all, the *true* data-generation process may actually *be* a higher-order polynomial instead of a linear one as in Figure 27.) But the more complications are necessary, the less plausible the explanation becomes (as long as the number of observations explained and the inherent randomness in the data remain constant). At some point, the explanation may be replaced by an alternative theory that strikes a better balance

[102] The process that actually generated the data in both Figure 27 and Figure 29 happens to be linear, $y = a + b\,x$, with two parameters a and b. The observed values of y are perturbed by random error (e.g., imperfect measurement), and as a result, do not fall exactly on the best-fitting straight line, as seen on the left of Figure 27. The cognizer, who observes only the pairs of x-y values in Figure 27, would like to find a formula that explains the observed data and that will enable her to predict future y values from observation of x values. Figure 27 shows two models (y1, y5) that she might use to do this, which vary inversely in simplicity and fit to the original data set. Figure 29 shows what happens when she applies each of these models to *new* data, generated from the same process (linear function with the same slope, intercept, and error generating process, but different specific random errors). The complex curve on the right (y5), which is partly based on random aspects of the original data, naturally fails to capture the independent random aspects of the new data and thus fares poorly. The simpler curve on the left (y1), which was poorer in fitting the original data set, provides a better fit to the new data because it ignores the random aspects of both data sets. Forster and Sober show how Akaike's theorems can be used, in conjunction with certain assumptions, to precisely quantify the tradeoff between fit and number of adjustable parameters. This is, in fact, the basic rationale for using significance tests for components in a multiple regression.

between simplicity and fit to current data and as a result is likely to be a more reliable guide to the future.

Measures that involve length and number of processing steps render simplicity *path-dependent*: The present ranking of theories by simplicity will depend not just on the syntax or semantics of the theories itself, but on previous linguistic and non-linguistic experiences of the cognizers who use the theory. Accidental events, in addition to than intrinsic features of the theories themselves, will affect theory choice. This is neither surprising nor objectionable as long as all that is at stake is pragmatic utility. It is a different story, however, if we wish to claim that simplicity increases accuracy or the chance of truth. In that case, we would want to show that a large variety of different historical paths are likely to converge on reasonable approximations to an objectively accurate solution. Akaike's measure of complexity of a model is in fact language-independent: the number of different curves (which is related to the number of adjustable parameters) that are possible before using the current data to estimate the parameters (Forster, 1999b). However, path-dependence enters the picture through the influence of prior knowledge on Akaike's measure.

The number of adjustable parameters in a model depends on background beliefs that are not themselves under evaluation. Forster and Sober (1994; Forster, 1999a, 2001) emphasize that only *adjustable* parameters count against simplicity, not parameters that have been adjusted independently or fixed by prior information. That is, if two theories have the same fit to a specific set of data, the theory that uses the greater number of parameters *to achieve that fit* is likely to more accurately predict future data. Even if the winning theory requires a large number of parameters or auxiliary hypotheses, if they are supported by independent evidence or background knowledge and are not tailored to the current data, they do not detract from the fit achieved by that theory to the current data. Previously fixed hypotheses or parameter values do not add complexity because the beliefs that support them are, at least temporarily, not open to question or change. To maximize simplicity of a specific mental model or theory, it should introduce as little as possible that is new, i.e., it should be based as much as possible on already established beliefs. This feature is so important that it has been regarded as an explanatory virtue in its own right: *conservatism.* As a result, estimates of simplicity will depend on what has previously been learned and on the questions the cognizer chooses to ask in the inquiry.

The effects of path-dependence can be mitigated to some degree by flexibility in choosing the scope of inquiry. If a simple and comprehensive theory cannot be found within the initial, constrained context, it is always possible and sometimes necessary to shift the focus, and thus expand the scope, of the inquiry to bring former background beliefs under scrutiny. At the same time, the experiences upon which those beliefs were based will also become relevant in the evaluation, and Akaike's theorems apply to the enlarged investigation. Thus, as the scope of the inquiry expands, a larger number of adjustable parameters or auxiliary hypotheses will have been balanced against fit to an expanded body of data. Modularity induced by causal structure offers promise that sequential strategies of belief evaluation will produce reasonably accurate results. A sequence of inquiries might gradually expand the scope of questioning by shifting focus backwards along causal links. Adjustments in increasingly general hypotheses may be investigated until the cognizer is able to find a simple and comprehensive solution. As the inquiry expands in scope, however, a price will be paid in terms of the number of resulting

changes in beliefs. Conservatism will be compromised as the cognizer poses a larger number of questions.[103]

Even as the scope of inquiry expands, evaluations of comprehensiveness and simplicity always take place within a context of background beliefs and evidence that are, *for the time being*, accepted as given. There is no way to step outside the belief system altogether and evaluate it as a whole. Among the background beliefs that usually (but not always) stay in place must be included some that pertain to the standards of simplicity itself. Simplicity is not an inherent or self-evident property of a set of beliefs. The degree of simplicity of a particular set of beliefs typically depends on non-obvious, contingent aspects of the domain and the context.[104] Processing strategies that are sensitive to such properties will be more likely to produce simple theories:

1. *Scope of inquiry*. Both comprehensiveness and simplicity must be evaluated relative to an (at least temporarily) unquestioned part of the belief system. They thus depend on decisions regarding the scope of inquiry, which determine which questions are relevant and which are not. Those decisions in turn depend on the chance of successful completion of the inquiry within current constraints balanced against costs of expanding its scope by shifting attention to a broader range of beliefs.
2. *Level of detail.* Forster and Sober's strategy involves empirical hypotheses or assumptions about the size of the random (i.e., unpredictable) component in the occurrence of events in the relevant domain. Decisions about the most desirable degree of resolution of an explanatory account depend on estimates of the inherent predictability of that particular domain. (These in turn might be based on past records of successful prediction in the domain in question.) The larger the signal-to-noise ratio, the more questions may be asked and the more detailed the resulting account might be.
3. *Tradeoffs among types of errors.* The choice between a simpler and a more complex theory may involve tradeoffs that balance costs and risks of different types of errors. Roughly speaking, adopting a more complex model improves the probability of a *hit*, but it also increases the *false alarm* rate (cf. Lehrer, 2000). For example, in Akaike's framework, the payoffs for correctly predicting details not captured by a simpler model are balanced against the penalties for supposing such deviations exist when they are due to chance (Figure 29). As Lehrer points out, the aim of theory selection is not to maximize probability of truth; that is easily accomplished by believing nothing except logical or mathematical theorems. The aim instead is to maximize overall reliability or utility, i.e., to balance the risk of accepting a falsehood against the risk of missing something true through excess of caution. Context-specific

[103] If the inquiry were somehow expanded to evaluate the belief system as a whole, *all* parameters would, in principle, be adjustable, all auxiliary hypotheses would be subject to revision, and all (memories of) observations ever made by the cognizer could be cited as evidence. But evaluation on this broad a scope is not only not computationally feasible, it makes no sense, as discussed in the next paragraph.

[104] By contrast, according to foundationalism, whether or not a belief is basic is a self-evident property of that belief.

objectives will influence the way cognizers make these tradeoffs, and will thus play a role in judgments of simplicity and acceptance of theories.[105]

4. *Weights on types of simplicity.* The multi-dimensional character of simplicity introduces uncertainty regarding the best balance among its different components.[106] One aspect of simplicity appears to be conformity to a pre-existing vocabulary of *concepts* (see footnote 98), which contributes to pragmatic utility because it provides more concise (i.e., simpler) descriptions. As noted in footnote 98, this advantage is relative to experience and language, since different concepts would be more concise in different languages and more efficiently represented by those with extensive experience with the concept in question. Another dimension of simplicity, which is also relative to a language and experience, is measured by the number and difficulty of processing steps required to acquire the theory or to apply the theory. Do these and other dimensions of simplicity have *language-independent empirical* advantages beyond pragmatic utility? According to McAllister (1996) there might be empirical significance in simplicity if tradeoffs among different dimensions of simplicity are calibrated by actual success and failure of theories. Implicit hypotheses about the predictive reliability of theories are represented by weights placed on different dimensions of simplicity. Cognizers would use simplicity as a filter that determines the types of theories or models they seriously entertain. Hypotheses about simplicity are then retained or rejected based on the actual success of the theories embodying the different types of simplicity.

The explanatory coherence of a set of beliefs thus depends on the scope of inquiry, the appropriate degree of resolution, trading off the chance of missing real patterns against the chance of inferring illusory ones, and assigning weights to different dimensions of simplicity. All of these decisions are essentially pragmatic, depending on factors like the cost of inquiry and its potential benefits, the statistical accuracy of prediction in a domain, the payoffs and costs of the current task, and the historical association of different types of simplicity with success. The reliably of the belief-generating process of the cognizer depends on whether it has operated with correctly set parameters. It is asking too much to suppose that cognizers consider all these factors explicitly in order to have justified beliefs. More often, the factors influence reasoning via

---

[105] The probability of error can be minimized simply be asserting nothing. Similarly, a single proposition is less likely to involve error (and thus, is more probable to be true) than a conjunction of two propositions, but two propositions may be more *useful* than one if each captures a regularity in the data. The desirability of taking a risk with one's beliefs (in Levi's (1986) phrase, *Gambling with the truth*) may account for the so-called conjunction fallacy (Kahneman, Slovic, & Tversky, 1982), in which people assess the "probability" of a conjunction (e.g., *feminist lawyer*) as higher than the "probability" of a conjunct (e.g., *lawyer*). The conjunction may be more useful for prediction in realistic contexts because it is less likely to overlook significant events, even though it is literally less likely to be true.

[106] For example, Forster & Sober (1994) admit that number of parameters is not the only dimension of simplicity that influences curve selection. A highly irregular curve might be invented that nonetheless had few or no free parameters. Such a curve would be "simple" in the Akaike framework, but would appear "complex" to humans. Akaike's theorems show that curves with fewer free parameters (no matter how complex they are in other respects) will tend to predict new data better than curves with more free parameters and the *same fit* to the current data. But he does not show that such curves will predict new data *well* (Swinburne, 1977). Similarly, an elaborate paranoid fantasy might have no auxiliary hypotheses (because it is held unchanged in the face of current data), but would also be judged complex. In other words, the Akaike framework assumes a pre-existing vocabulary of "simple" – i.e., familiar – theoretical concepts and principles (including curves).

inherited traditions in specific domains or disciplines and/or the specific experiences of a cognizer in a domain. If they are made explicitly, judgments about such factors are made from a third-person, externalist point of view. In dialogue terms, it is the responsibility of the judge (or of the cognizer acting as judge) to determine, for example, if a mental model or story has "overfit" the data. It is the judge's job to terminate a dialogue if details have been elaborated beyond what the experience and knowledge of the dialogue participants (or cognizers) warrant. Coherentism began as an internalist theory like foundationalism. But efforts to clarify the criteria for coherence show that an externalist component is an essential part.

We observed earlier that comprehensiveness of explanation is an empty achievement unless the explanation is also simple. It is equally true that explanatory simplicity is meaningless without comprehensiveness. Conciseness can be maximized and computational effort can be minimized by suspending *all* beliefs (including estimates of parameter values and auxiliary hypotheses), and thus explaining nothing. For simplicity to be a meaningful evaluative measure, it must be subject to the countervailing pressure of comprehensiveness: A theory must explain as much as possible (comprehensiveness) using the most economical possible means (simplicity). The explanationist interpretation of coherentism thus implies a distinction between the *roles* of explaining and being explained. Simplicity focuses on the conciseness of the beliefs that do the explaining and the efficiency with which predictions can be generated from them, while comprehensiveness focuses on the number, variety, and precision of the beliefs that are to be explained (Figure 25). The to-be-explained beliefs typically include reports of perceptual observations, but such beliefs are not "basic" in the foundationalist sense. In the coherentist framework, explanation and explananda provide mutual support for one another and are accepted or rejected as a package. Competing explanations may explain the same data differently, but they may also differ in the data that they explain and in how events are matched up to potential causes. The best overall explanatory account may involve dismissing some of the data as anomalous and providing different types of explanations for different subsets of the data. Evaluation by simplicity and evaluation by comprehensiveness must proceed in parallel, and apply simultaneously to *both* the explanation *and* what it explains. Because coherence comprises both simplicity and comprehensiveness, it captures an essential aspect of good explanations.[107]

How Do We Know a Hypothesis Is Correct?

Testability

A traditional sign of a good explanation, inherited from foundationalist thinking, is that the individual hypotheses within the explanation are testable. An individual hypothesis is testable if there are observations that would cause the cognizer to give that hypothesis up (Quine & Ullian, 1970, p. 50) – that is, if there are predictions whose violation would disconfirm the hypothesis. While comprehensiveness and simplicity fit nicely with the coherentist concept of mutual support, testability at first appears not to. The holistic aspect of coherentism seems in fact to *preclude* testability for *individual* explanatory hypotheses. Closer examination, however,

---

[107] Informal logicians and contemporary foundationalists interpret explanation in terms of linear argument, e.g., as a form of *abductive* inference of an explanation from a *given* set of data. The problem with this is that the data are not in fact "given," but in the real world are selected, evaluated, classified, and interpreted so as to produce the overall best combination of theory and data. Abductive inference should properly be interpreted as the outcome of this coherence-improving process.

shows that holism (properly understood) yields a more powerful and more sophisticated approach to testability.

In forming a hypothesis (e.g., about the enemy's objectives), our intent is often to anticipate future events. A single hypothesis, however, is usually not enough. It takes additional hypotheses (e.g., regarding enemy doctrine, order of battle, equipment, weather, and terrain, as well as innumerable implicit common sense beliefs) to make a prediction (e.g., where the enemy might deploy its troops and where it might attack). *Holism* is the view that mutual support among many beliefs is usually necessary to generate and justify predictions. It follows that failure of those predictions does not, by itself, point to any single belief as the culprit. Holism thus implies that beliefs must be tested as a body rather than individually. If observations run counter to expectations, they contradict what is in effect a conjunction of beliefs, but do not specifically disconfirm any particular proposition within that conjunction. A conflicting belief, by itself, means that a change *somewhere* in the relevant set of beliefs should be made, but does not tell us *where*. [108]

It seems to follow from holism that individual hypotheses are not testable. A particular explanatory hypothesis (e.g., that the enemy will attack in the north) can always be saved from apparent refutation (e.g., the observation that there is no artillery in the north) by revising one of the other hypotheses (e.g., by assuming that the enemy have longer range artillery, or that artillery in the north was missed due to concealment). As Quine (1953, p. 43) says, "Any statement can be held true come what may if we make drastic enough adjustments elsewhere in the system." The mere existence of sound arguments on both sides of an issue, or of more than one consistent theory, is taken, falsely, to show that there are no differences in merit among the accounts (Laudan, 1998). [109]

---

[108] Quine (1953, p. 41) criticizes simplistic ideas about testability as follows: "The dogma of reductionism survives in the supposition that each statement, taken in isolation from its fellows, can admit of confirmation or infirmation at all. My countersuggestion…is that our statements about the external world face the tribunal of sense experience not individually but only as a corporate body." Quine explicitly labels this view *holism* (Quine, 1990, p. 10). Duhem (reprinted 1998) had earlier made these points in regard to physics, but they also apply to other areas of thought (Gillies, 1998). The impossibility of crucial experiments that directly disconfirm individual theoretical hypotheses is called the Duhem-Quine thesis in philosophy of science (Curd & Cover, 1998). We suggest that it also applies to many common sense "theories," e.g., about enemy intent, that are considered in tactical military settings.

As we saw earlier, holism is one of the reasons for rejecting foundationalism in favor of coherentism (the other reason was the difficulty of making sense of the distinction between basic and derived beliefs). Holism is simply a restatement of the fact that linear argumentation cannot generally resolve conflicts in evidence. Just as there may be more than one consistent account, so there may be more than one "sound" argument for the revision of some beliefs based on other beliefs that continue to be accepted. Figure 21 and Figure 22 illustrate this. If we accept the truth of a hypothesis (the enemy will attack in the north) based on certain reasons, but then run into evidence against the hypothesis (no artillery in the north), we automatically have an argument for any defeater (e.g., longer range artillery) that undermines the conflicting evidence. Conversely, if we encounter the conflicting evidence first and therefore accept the negation of the hypothesis (the enemy will attack in the south), we automatically have an argument to "defeat" any of the reasons in favor of the hypothesis. Arguments that would be individually judged reasonable can lead to contradictory conclusions. Thus, overall accounts must be compared.

[109] It is sometimes a meaningful *option* to save one hypothesis by modifying others. As a result one, when testing a hypothesis, one is sometimes (at least implicitly) thinking about a body of hypotheses – i.e., making decisions to accept or reject other, related hypotheses as well as the hypothesis in question. Thus, holism asserts merely that in *deciding* whether to accept or reject an individual hypothesis, other, mutually supporting hypotheses must often be taken into account. It does not follow, however, that saving a hypothesis by modifying others is always an *acceptable* option. For example, it may incur an exorbitant cost in complexity or excessively disrupt our pre-existing

Coherentism did not create this problem; on the contrary, it is motivated by it and offers a solution to it: Reformulate the testability criterion in coherentist terms. Holism does imply that individual beliefs cannot be evaluated by themselves, and thus are not accepted or rejected on the basis of isolated arguments. But they *can* be evaluated – albeit indirectly through their membership in bodies of beliefs, which are in turn evaluated *as a whole* with respect to explanatory virtues. An individual hypothesis can be tested by comparing a body of beliefs *containing that hypothesis* to an alternative body of beliefs *that does not contain it*. The observations disconfirm an individual hypothesis if the most coherent body of beliefs containing that hypothesis is less coherent than the most coherent body of beliefs not containing that hypothesis. Thus, whether a particular hypothesis is testable depends on whether there is a clear difference in coherence between accounts that do and do not contain that hypothesis. This points to a dialogue-based method for resolving conflicting evidence. To evaluate a hypothesis in the light of observations, the cognizer should first act as proponent of the hypothesis, constructing a story that contains the relevant observations, the hypothesis, and other assumptions in order to make the combination as coherent as possible (i.e., simple, comprehensive, and minimally disruptive of pre-existing beliefs). Then, acting as an opponent of the hypothesis, the cognizer should construct another story containing the relevant observations, the negation of the hypothesis, and other assumptions to make *this* combination as coherent as possible. The observations disconfirm the hypothesis if the latter story is more coherent than the former, within

---

beliefs. Duhem (1998, pp. 277-278) was well aware of this and emphasizes the role of "good sense" in addition to logic in deciding among theories.

Quine, however, has famously argued that theories are underdetermined by all possible evidence (Quine, 1990), i.e., that there exist theories that are incompatible with one another but which it would be equally reasonable to accept. Since they are incompatible, there must be at least one statement asserted by one theory and denied by the other (but Quine, 1990, pp. 13-14). Underdetermination therefore implies that some individual hypotheses are not testable. But as Quine himself sometimes acknowledged (Quine, 1990), this claim goes well beyond the Duhem-Quine thesis. It is far from being essential to the coherentist position, and Quine's defense of it is confusing and ultimately unconvincing. Quine sometimes describes the Duhem-Quine thesis as though it were equivalent to underdetermination (hence, non-testability of individual hypotheses). For example, Quine says that any statement can be saved from disconfirming evidence by making changes elsewhere in the theory. This is nothing more than plausible holism if taken to describe an *option* that is sometimes meaningful and ought to be considered. But if it is taken to (always) describe an *acceptable* option, it is equivalent to a very broad and implausible claim about underdetermination of theories. It would follow that for *any* statement, there is an acceptable theory containing that statement and also an acceptable alternative theory containing the negation of that statement (Laudan, 1998).

Quine may see these two interpretations as equivalent because he believes that any theory that is both logically consistent and fits the data is as good as any other. Quine arrives at this view by denying that simplicity and conservatism have anything more than pragmatic value. If they have no empirical significance, they cannot be appealed to resolve underdetermination. This view does not do justice to the practices of reasoning and justification of beliefs. As we saw previously, cognizers take simplicity seriously enough that it often *outweighs* fit to current observations and strongly influences empirical predictions. Quine himself is not consistent on this. Quine and Ullian (1970) not only propose testability as a criterion of theory choice (p. 50); they acknowledge the empirical relevance of simplicity by endorsing an evolutionary account, according to which simplicity criteria survive insofar as they support predictive success (p. 47). (This is similar to McAllister's account, according to which weights on different simplicity dimensions are adjusted in response to empirical successes and failures of the theories that score high or low on those dimensions.) Underdetermination is far less likely (but not ruled out) if simplicity and conservatism combine with comprehensiveness and logical consistency to form the "good sense" which Duhem said was crucial in theory evaluation.

the constraints laid down by the judge on scope of inquiry, level of detail, relative costs of errors, and relative importance of different simplicity dimensions. [110]

We do not need to chose between holism and testability. Unless there are some circumstances under which a belief would be given up, it contributes nothing to the usefulness of the explanation as a whole, i.e., its ability to predict. On the contrary, to understand testability, we need to understand how individual beliefs contribute, or fail to contribute, to the generation of predictions by bodies of belief. Holism is not incompatible the testability of individual beliefs – as long as explanatory virtues are sufficient to discriminate among explanatory accounts that do and do not contain the individual beliefs in question. Simplicity and comprehensiveness do appear to promote testability. The more comprehensive a theory tries to be (for example, the more different types of predictions it makes and the more precise they are), the more opportunity there will be to find discrepancies between observational beliefs and theory. At the same time, the fewer auxiliary hypotheses or adjustable parameters a theory draws on, the less room there is for saving a hypothesis from disconfirming evidence by making changes elsewhere in the web of beliefs. Simplicity narrows the range of feasible responses to disconfirming evidence, and increases the chance that the hypothesis in question will be rejected in the face of unexpected data. So, comprehensiveness and simplicity will lead us to prefer theories in which individual beliefs tend to be more testable.

Simplicity, comprehensiveness, and logical consistency by themselves are far from being sufficient for testability. From the pure coherentist perspective, theories are not tethered to anything outside themselves. Thus, nothing prevents the invention of new bodies of belief, complete with both explanatory hypotheses and tailor-made "observation beliefs." Such fictional constructions might clearly be simpler and more comprehensive than current theories. Just as there is no limit to the number of fictional plots, so there is no limit to the number of coherent theories that might be devised as long as the theorist is free to invent the "data" to be explained.

---

[110] Quine and Duhem also differed on how wide a net holism casts – i.e., the size of the bodies of belief that are evaluated. Duhem took a narrow view of the scope of belief revision: A small fixed group of hypotheses generates a prediction and only the beliefs in that group are candidates for revision in explaining a failed prediction. Quine points out that in principle at least, the entire system of beliefs is what generated the prediction, and a failed prediction generates an adjustment process that might reverberate throughout the belief system. The two views can be reconciled (cf., Gillies, 1998) by recognizing the iterative character of belief revision in the face of conflicting evidence. Cognizers will begin by trying out revisions in a small set of closely relevant hypotheses. If adjustments to beliefs within the initial mental model are unsuccessful in resolving conflict, the scope of inquiry may widen to include examination of more distantly related hypotheses. The horizon of inquiry will never, however, encompass the entire belief system.

Scope refers to the range of questions that are asked; underdetermination refers to the uniqueness of the answers. Both underdetermination of theories and the scope of belief revision are best regarded as empirical questions, to be answered in part at least by study of reasoning in the relevant domains rather than by general arguments (cf. Feynman, 1965, and Weinberg, 1992). The scope of belief revision depends on the actual extent of the causal structure and inferential connections among disparate elements in a belief system. Underdetermination of theories depends on the availability and diagnostic power of empirical evidence in a domain and on the tendency in that domain for different dimensions of coherence to offset one another and prevent a clear winner. Other things being equal, a wider scope of belief revision might be positively correlated with underdetermination because in both cases there would be more possible combinations of beliefs available. On the other hand, a wider scope might in fact be associated with less underdetermination if cognizers expand the scope of inquiry (e.g., make a new distinction) in order to *resolve* a tie between two theories. (Thus, contrary to some Quine interpreters, wide holism neither follows from nor implies a view on underdetermination of theories.)

Constraints on freedom of invention are necessary to anchor coherentism in realty. This is the role sometimes assigned to an additional criterion for explanatory goodness, called *conservatism.*

Conservatism

There is always more than one *possible* way to handle conflicting evidence. One strategy is to start from scratch and simply manufacture more easily explained "data." If this solution seems unjustified, so do a variety of less drastic choices that comprehensiveness and simplicity alone appear to permit. For example, a cognizer might explain a single unexpected observation (e.g., of artillery placement) by questioning specific assumptions (e.g., about artillery range or concealment). Alternatively, she might explain the observation by revising general assumptions about the enemy's overall strategic objectives and doctrine. Even though both approaches would eliminate the conflict, the latter is more sensible at least at the start. In short, conservatism in the revision of beliefs is a virtue. Conservatism implies that in reducing conflict, we should first look for explanations that minimally disrupt the beliefs we already have, and only if that fails, consider explanations that overturn fundamental assumptions and may have ramifications throughout our belief system. Other things being equal, in seeking to improve our understanding of the world, we should minimize the total change that results either by adding new beliefs or dropping old ones (Harman, 1973, p. 159; Lycan, 1996, p. 5).

Conservatism applies, in the spirit of coherentism, to all beliefs without distinction, not to a restricted class of basic beliefs as in foundationalism. Conservatism implies that "the bare fact of one's holding a belief renders that belief justified, to some degree; any belief at all is at least minimally warranted" (Lycan, 1988, p. 162).[111] On the face of it, this turns critical thinking on its head. Critical thinking began in a quite different, rationalist spirit, insisting that we can and should *doubt* all our inherited beliefs. (It then prescribed a process of carefully re-acquiring beliefs one at a time by methods that establish absolute certainty.) Conservatism insists, on the contrary, that doubt is irrational unless there is a specific *reason to doubt*. It thus shifts the burden of proof, from the demand that every belief be justified to the demand that every *change* in belief be justified (Harman, 1999a, p. 27). As Lycan (1988, p. 162) points out, conservatism keeps us from changing our beliefs for *no reason at all*. Foundationalism corresponds to a dialogue in which there is a fixed starting point (basic beliefs) and every new assertion must be defended. Coherentism corresponds to a dialogue in which the participants can begin from any mutually agreed upon point and need not defend or revise an assertion unless it is specifically challenged.

Like simplicity, conservatism has a clear pragmatic rationale. In some respects, in fact, the pragmatic impact of conservatism overlaps with that of simplicity. For example, we have already seen how the use of familiar concepts and beliefs contributes to the simplicity of a theory, measured as its conciseness, and how the familiarity of processes reduces the computational effort required to apply a theory. Nevertheless, simplicity and conservatism often do not agree: Abandonment of familiar concepts and restructuring of theory might in some cases lead to a radically simpler theory. Conservatism will nonetheless suffer, because it also includes the contribution of familiarity to ease of learning the theory. The more a viewpoint disrupts

[111] Harman (1999b) characterizes his own view as "general foundationalism." According to Harman, "A special foundations theory holds that only certain specified beliefs and inferential procedures are foundational. A general foundations theory holds that all of one's beliefs and inferential procedures at a given time are foundational at that time." That is, they are accepted in the absence of reasons to the contrary.

previous beliefs, the more difficult it will be to learn the viewpoint and to communicate it to others.

The first and most basic rationale for conservatism is that it is a prerequisite for *reasoned* change of belief. Improvements in simplicity and comprehensiveness naturally lead us to revise our understanding of events, and may even overturn fundamental assumptions. But to be justified, such improvements must outweigh the penalty in conservatism that that is incurred whenever we jettison previously held beliefs or add new ones. The more drastic the revision in our beliefs, the greater must be the gain in simplicity and/or comprehensiveness.

Quine (1953, p. 79; 1960, pp. 3-4) cites Neurath, who compared our system of beliefs to a boat which we must repair while it keeps us afloat. In order to pull out rotten planks in one part of the boat, we must stand on another part. We may ultimately replace all the planks, but we had better do so in stages. Similarly, to challenge a belief, other beliefs must be taken for granted to serve as reasons for the challenge, even though those beliefs may themselves ultimately come under challenge. The more of our current beliefs we keep intact, the more persuasive a case we can make for any particular coherence-increasing step. Conservatism maximizes coherence with, and plausibility in the light of, our *current* beliefs, while simplicity and comprehensiveness generate pressure for change.

Conservatism thus imposes a sequential character on evaluation of beliefs. We take bite-size chunks of the belief system for consideration in the light of other beliefs, while most of our beliefs stay (temporarily at least) in the background. A sequential strategy of precisely this kind is necessary in any case to deal with computational limitations that prevent simultaneous evaluation of the large numbers of beliefs. The two strategies discussed earlier – evaluation of mental models and a limited horizon of spreading activation – both embed conservatism. Pragmatic factors therefore constitute a second rationale for conservatism.

A third rationale for conservatism arises from a limitation of coherentism: its inability to explain how justification of beliefs "gets started." Comprehensiveness and simplicity do not kick in until a "system" of beliefs is already in place, including both explanations and explananda. Perceptual and memory beliefs, however, can be justified by conservatism before they begin to cohere with other beliefs in a larger system. In other words, conservatism accounts for justification in the absence of "reasons" – in particular, when there are no inferential links to other beliefs.[112]

This strength, however, seems to lead to problems: Conservatism applies not only to legitimate perceptual and memory-based beliefs, but to any beliefs at all. The only requirement is that we already believe them. Conservatism does not rule out alternative absurd systems of beliefs, which would be justified if we in fact believed them. To some, such as Lehrer, conservatism also imposes an undue and arbitrary constraint on improvement of our belief system. In the next section, we will explore an alternative to conservatism based on an externalist, their-person point of view.

---

[112] Conservatism is not the only possible solution to the problem of isolated beliefs. One might prefer to say that isolated beliefs are not in fact justified. Even more strongly, one might say that *beliefs* cannot exist in the absence of systematic relations to other beliefs. (In these cases, one would have to say that very young children and animals have no beliefs or no justified beliefs.)

Why Is Perception Special?

Even if no beliefs are known with certainty, some beliefs clearly do receive priority over others. In particular, perceptually based beliefs about nearby objects in plain sight (as well as recollections of recent perceptual experiences) have high degrees of credibility even when they are not highly interconnected with other beliefs. Coherentism cannot easily explain where this credibility comes from.[113] Since pure coherentism does not acknowledge any grounding of beliefs in sense experience (or beliefs about sense experience), it provides no basis for rejecting an absurd system of beliefs that happens to be internally coherent. This leaves the door open for skepticism.[114]

At least three solutions have been proposed for this problem by coherentists:

1. Reflective beliefs about perception
2. Conservatism
3. Independent weight on perceptual beliefs

The first solution is internalist and intellectualistic, while the third, if it has any rationale at all, marks a dramatic break with internalism.

Coherence with reflective beliefs about perception. One response is to maintain that perceptual judgments are especially coherent with the rest of one's beliefs as a result of reflection (Bonjour, 1985; Lehrer, 2000, pp. 138-144; Harman, 1973). The priority of perceptually based beliefs is a consequence of their coherence with meta-beliefs about how reliable perception is under relevant conditions. This solution relies heavily on internalist assumptions. It requires that the cognizer be explicitly aware of all her beliefs, and that in addition she have second-order beliefs that explain the first-order beliefs. The explanatory beliefs concern the causal origins of the first-order beliefs (e.g., from perception, memory, logical intuition, or reasoning) and express judgments about the reliability of beliefs that have those origins.

This proposal has several defects: First, it saves coherentism by significantly amending it. It constitutes an admission that pure coherence is insufficient to support the justification of

---

[113] There is an enormous amount of coherence among perceptual beliefs over small intervals of time, and this certainly affects their credibility. But the degree of overall coherence with other, non-perceptual beliefs of the cognizer may be low. It is tempting to speak of the comprehensiveness criterion as if it referred to explanatory coverage of *data*. But nothing is given or marked off as data in a coherentist framework. When an observational belief conflicts with an explanatory belief, coherence does not prevent us from rejecting the observational belief (i.e., elements in the bottom row of Figure 25) instead of the explanatory belief (i.e., an element in the top row of Figure 25). Moreover, nothing prevents us from accepting a large number of entirely fictitious observational beliefs along with a simple explanatory hypothesis that accounts for them. The resulting system might score very high on simplicity and comprehensiveness, but does not seem justified.

[114] Coherentism is an attempt to handle type C challenges, i.e., the existence of competing points of view. Pessimism about resolving such challenges is precisely what motivated skepticism historically. According to one of the earliest skeptical philosophers, Sextus Empiricus (1990, p. 24): "…the skeptic found himself involved in contradictions of equal weight, and being unable to decide between them suspended judgment…" Critics of coherentism claim that it fares no better than skepticism. According to C. I. Lewis (quoted by Firth in Troyer, 1998, p. 206): coherentism '…strikes me as supposing that if enough probabilities can be got to lean against one another they can all be made to stand up… I think the whole system of such could provide no better assurance of anything in it than that which attaches to the contents of a well written novel."

beliefs. Specific *kinds* of beliefs (i.e., second-order beliefs about reliability) must be part of the mix, with which other beliefs must cohere. This is a new form of foundationalism, in which second-order beliefs about the reliability of belief-generating systems are treated as basic (instead of the beliefs actually delivered by those systems). Second, an equally coherent system might be constructed that includes second-order beliefs about the reliability of fictitious belief-generating systems and thus rejects the perceptual beliefs that we currently regard as reliable. To rule this out, the proposal would have to stipulate not just the form but also the content of the second-order "basic" beliefs. This brings the theory even closer to foundationalism. Finally, the demand for continuous reflective awareness exceeds human capabilities, implausibly denies justification when higher-order reflection is absent, and threatens a vicious regress involving beliefs about beliefs, beliefs about those beliefs, and so on (Sosa, 1991, pp. 205-207).

Conservatism. A second approach is to supplement coherence by a principle that enjoins us to hold on to the beliefs we have. Conservatism is a less intellectualist way to favor perceptual beliefs in a coherentist framework (e.g., Harman, Lycan, Quine & Ullian). As already noted, conservatism tells us that, other things being equal, it is preferable to minimize the total change that arises from adding new beliefs or dropping old ones (Harman (1973, p. 159). By virtue of conservatism, "the bare fact of one's holding a belief renders that belief justified, to some degree; any belief at all is at least minimally warranted" (Lycan, 1988, p. 162). In particular, any "spontaneous" beliefs (e.g., from perception or memory or anywhere else) that we find ourselves holding cannot be jettisoned without cost (Lycan, 1996, p. 5). Conservatism fits well with coherentism because it applies to all beliefs without distinction, not just to a restricted class of basic beliefs as in foundationalism. Conservatism also explains how belief systems "get started." Perceptual and memory beliefs, for example, can be justified by conservatism before there is a belief *system* that can be meaningfully evaluated in terms of coherence.

As we have already noted, a problem with conservatism as a solution to the problem of perceptual grounding is that it is not specific enough. The support it provides for perceptual beliefs is the same as the justification it provides for all other beliefs. Conservatism will be equally satisfied if one decides to keep a hypothesis and reject an observation that conflicts with it, rather than the other way around. In fact, the canons of minimal belief revision proposed by Gardenfors (1992, p. 17) seem to favor giving up an observation in preference to giving up an explanatory hypothesis that it supports. The canons imply that grounds should be given up before conclusions. If grounds are rejected, a conclusion may be retained without inconsistency; but if the conclusion is rejected, at least one of its grounds must also be rejected in order to avoid a contradiction (i.e., believing the grounds and the negation of the conclusion). Rejecting data thus leads to less extensive change in the overall system.

Cognizers in fact do tend to start out, as children, with beliefs based on perception and memory rather than more elaborate inferences (Lycan, 1996, p. 19). Thus, as a matter of fact conservatism might tend to give priority to these beliefs simply because they come first. But one could imagine an abnormal developmental process in which a cognizer was deprived of sensory stimulation and developed hallucinatory beliefs instead of the normal perceptual ones. According to conservatism, such beliefs would be just as justified as ordinary perceptual beliefs.

Conservatism fails to provide a convincing account of why perceptual and memory beliefs receive priority.[115]

Although conservatism is often proposed as a component of coherentism, and coherentism is supposed to be an internalist theory. Nevertheless, the role of conservatism in helping belief systems "get started" implies that it is not an internalist criterion. Internalism requires that all the factors relevant to justification be consciously known. But before a particular age, children are not likely to have reflective beliefs about what they believe (e.g., Kuhn, Amsel, & O'Loughlin, 1988). In order to even pose questions such as, *Do I really believe that, and why?*, they would already have to possess a sophisticated belief system, and there would be no need to use conservatism as a method for getting started. If their beliefs are nevertheless justified to a degree simply by virtue of being believed, then cognizers do not consciously grasp all the factors relevant to justification (in particular, the fact that they have certain beliefs). Having opened the door to externalist criteria, at least implicitly, it might be worthwhile to look for some more effective externalist method for grounding beliefs in perception and memory.

Independent weight on perceptual beliefs. A third approach to grounding belief in perception is to single out the class of observational beliefs for special weight independent of their coherence with other beliefs. Haack (1993) proposes a synthesis of foundationalism and coherentism to account for the priority of perceptual beliefs. As we have seen, Haack compared mutual support among beliefs to the way intersecting words in a crossword puzzle constrain one another as they are filled in. But she extends the analogy to include *clues*, which correspond to observational data. Crossword puzzle clues represent additional constraints independent of those that words impose on other words that intersect them. Similarly, perceptual experiences supply a source of justification that originates outside the system of coherence relations among beliefs. Not every word has a clue; some words must be identified solely on the basis of constraints from other words. In the same way, some beliefs (e.g., inferences about enemy intent) are not directly supported by experience, but are related indirectly to other beliefs that are (e.g., about enemy movements and placement of artillery).

Thagard (2000, p. 43) applied Haack's idea to the design of a coherence-based inference engine. In Thagard's system, "propositions that describe the results of observations have a degree of acceptability on their own." Thagard specifies rules that associate an appropriate class of privileged inputs with every type of coherence (viz., explanatory, deductive, analogical, perceptual, conceptual, and deliberative). In each case, the independent weight given to such inputs can be overridden if the inputs are sufficiently incoherent with other elements.[116]

---

[115] The previous proposal, coherence with reflective beliefs, was able to single out beliefs produced in particular ways, e.g., by perception and memory, as especially reliable. The reflective position closest to conservatism, on the other hand, is a blanket belief in the trustworthiness of all the beliefs that a cognizer has accepted. Lehrer (2000, pp. 138-144) requires coherence with just such a higher order belief. This proposal combines the worst features of solutions based on reflection and conservatism. Besides the problem of not being specific enough, this analysis raises the issues of excessive intellectualism.

Harman (1986, pp. 57-59) suggests, as a canon for belief revision in addition to conservatism, that we not give something up that we can easily get back. Thus, we should not reject an observation if it can easily be repeated. This principle is easily by-passed, as Harman points out, if we simply explain the rejected observation by adopting a belief that there has been a (repeatable) observational error.

[116] Are these views foundationalist, coherentist, neither, or both? They reject the strict foundationalist claims that (i) there are beliefs whose justification is not affected by other beliefs, and (ii) all other beliefs receive justification *only*

These foundationalist-coherentist hybrids suffer from the same problems as traditional and contemporary foundationalism: How to explain the privileged status granted to certain classes of beliefs and not to others. The more appealing solution to the problem is to accept that some of the factors justifying belief acceptance may not be cognitively accessible, that is, to adopt a form of *externalism.* From the externalist point of view, the independent weight given to sensory beliefs is easily explained, in terms of the reliability of the relevant perceptual system under current conditions. Thus, if perceptual systems reliably produce true beliefs, then beliefs based on perception are justified, whether or not the cognizer reflects on how reliable they are (or how reliable her beliefs about their reliability are, and so on). Perception and memory may reliably anchor a system of beliefs in external reality without the subject's awareness. More generally, a belief, whether perceptual or inferential, may constitute genuine knowledge even though the cognizer is unable to articulate reasons for holding it.

In sum, coherentism is not without problems. They are, not surprisingly, the mirror images of the problems encountered by foundationalism:

1. If there are no basic beliefs, how can coherentism account for the priority given to perceptual beliefs?
2. How can bodies of beliefs be comparatively evaluated? Even for moderately sized belief systems, the combinatorics of coherence far exceed human cognitive capabilities (Cherniak, 1986). What are the criteria of coherence and how are they applied?

Both of these problems point the way to a dependency of coherentism on *externalist* assumptions. Unlike foundationalism, coherentism demands to be interpreted in externalist terms. We have already discussed computational feasibility and the requirements it generates for parallel automated constraint satisfaction mechanisms, for shifting attention among small mental models, and for the modularity implied by the causal structure of events. We have also discussed the role of externalism in determining the appropriate constraints on simplicity, such as scope of inquiry, degree of detail, tradeoffs among different types of errors, and weights on different dimensions of simplicity. Finally, we discussed how coherentism must add an externalist point of

---

by transmittal from the basic beliefs. But foundationalism might be defined more inclusively, as the recognition that there are two epistemically different classes of beliefs, and that *some* justification is transmitted from beliefs in one class to beliefs in the other. If so defined, the views in question count as a form of foundationalism. Similarly, these views reject the strict coherentist claim that all beliefs are on the same epistemological footing, and that all justification derives from coherence with other beliefs. But coherentism may also be defined more inclusively, as the recognition that all beliefs receive *some* justification from other beliefs. In that case, these views also count as a form of coherentism.

Quine has always been a dualist about justification, stressing the interaction of sensory stimulation and coherence among beliefs. According to Quine (1953, p. 42): "… total science is like a field of force whose boundary conditions are experience." According to Aldrich (quoted approvingly by Quine, 1960, p. 12): "…there are two forces that interpenetrate or fuse to constitute the field: the 'empirical' force extending into the field from 'outside' and thus being stronger near the periphery; and the formal or logical force, whose principle is simplicity and symmetry of laws, radiating out from the center." Thagard's device of associating intrinsic strength to specific observational beliefs is paralleled in Pearl's (1989) Bayesian causal network and in Shastri's parallel reflexive reasoning network (Shastri & Ajjanagadde, 1993; Cohen, Thompson, Adelman, Bresnick, Shastri, & Riedel, 2000). All three systems posit independent activation for observational facts. All three systems implement Quine's vision of a field of force that achieves an equilibrium encompassing both observation and coherence.

view to account for the priority given to perceptual and memory beliefs, in terms of the reliability of the relevant belief-generating processes.

10. RELIABILITY FROM A THIRD-PERSON PERSPECTIVE

According to Halpern (1996),

> "Critical thinking is the use of those cognitive skills or strategies that increase the probability of a desirable outcome. It is used to describe thinking that is purposeful, reasoned and goal directed - the kind of thinking involved in solving problems, formulating inferences, calculating likelihoods, and making decisions when the thinker is using skills that are thoughtful and effective for the particular context and type of thinking task. Critical thinking also involves evaluating the thinking process - the reasoning that went into the conclusion we've arrived at the kinds of factors considered in making a decision. Critical thinking is sometimes called directed thinking because it focuses on a desired outcome."

Many definitions of critical thinking refer to it as a deliberate or goal-directed process, but Halpern's definition is unusual in its emphasis on the likelihood of achieving an external outcome in a real context. Critical thinking is the "kind of thinking" that is likely to be "effective" in the relevant "type of thinking task" in the relevant context. This terminology refers to a population of cases in which successes and failures occur repeatedly over time, not a one-time process that need only be experienced as purposive "from the inside." This definition is an excellent example of a predominantly externalist approach to critical thinking.

Externalism has attracted considerable recent interest from philosophers (e.g., Goldman, 1992, 1986; Dretske, 1983; Nozick, 1981; Sosa, 1991; Plantinga, 1993b). According to one version of externalism, called *reliabilism* (Goldman, 1992, 1986), a belief is justified if it is generated or sustained by cognitive processes that reliably generate truths and avoid falsehoods under the relevant conditions. According to another type of externalism, called *virtue* or *faculty* epistemology, beliefs are justified if they are produced by properly functioning faculties operating correctly in the environments for which they were designed (Plantinga, 1993b). On both views, beliefs of different kinds are more or less justified depending on the processes and mechanisms that produced them and the specific conditions under which the processes were operating.

Externalism does not insist that people have cognitive access to reasons for a belief, that people have second-order beliefs about the reliability of first-order beliefs, or even that beliefs are always under voluntary control. A person, mechanism, faculty, or strategy is deemed successful or unsuccessful (expert or non-expert) based on performance and results, i.e., the actual accuracy of judgments under various conditions. Externalism accounts for our willingness to attribute knowledge to people even when they cannot accurately articulate the reasons for their judgments (Sternberg & Horvath, 1999; Berry & Dienes, 1993; Nisbett & Wilson, 1977). There is evidence that experts can become highly proficient in recognitional skills in which they are less able than novices to describe their own thought processes. For example, expert physicians are sometimes not able to retrieve the explanation supporting a diagnosis (Patel, Arocha, & Kaufman, 1999, p. 82). Externalism allows evaluation of a belief in terms of the objective effectiveness of strategies in the external environment, relatively automatic processes (such as perception, pattern recognition, and constraint satisfaction in connectionist networks), and features of cognitive mechanisms (such as processing capacity and the structure of knowledge in

long-term memory). It thus promises more fundamental integration with concerns of cognitive psychology.

The architecture of belief postulated by reliabilism is a series of input-output processes, as illustrated in Figure 30. Perceptual processes take environmental stimuli as inputs and produce beliefs as outputs. Memory and reasoning take beliefs as inputs and produce other beliefs as outputs. Some beliefs, of course, are produced by more than one process, either serially or in parallel. Such beliefs are justified if each process involved in their production reliably generates truths when true beliefs are given as inputs.

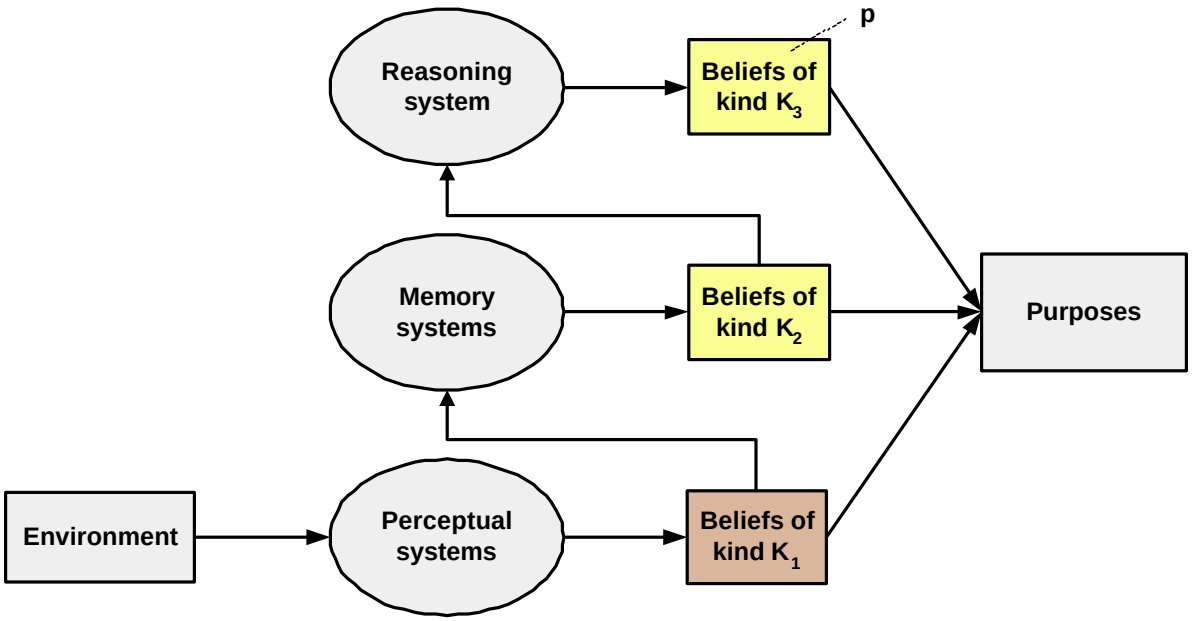


Figure 30. Reliabilist paradigm for acceptability of beliefs: a series of input/output processes. In the version shown here, inputs to perceptual faculties come from the environment, while inputs to memory and inferential faculties include the outputs of other belief-generation processes.

Both coherentism and reliabilism, in different ways, reject the central role assigned to *arguments* by foundationalism. They provide a new insight into how arguments are used to determine belief acceptability:

- For coherentism, arguments demonstrate inferential connections within a system of beliefs and thus help to show how tightly woven it is. Arguments by themselves do not justify beliefs. They justify the system as a whole by establishing its *overall* coherence compared to systems that lack the beliefs in question. Arguments illuminate the connections of *candidate* beliefs to other beliefs that are already within such a system. They may also show the need for revision of the system in order to accommodate new information. Acceptance of both arguments and of beliefs ultimately depends on how alternative systems are evaluated.
- Externalism places justification in a larger perspective, in which arguments are not necessary at all. Justification by reasons is one among many belief generation processes, like perception, recall, and pattern recognition. Each of these processes

must be evaluated for its reliability in producing true beliefs of specific types under specific circumstances. None of these processes has a privileged claim as a source of legitimate belief acceptance.

Despite its problems, some variant of foundationalism is by far the most popular approach in informal logic and critical thinking (cf., Freeman, 2000). With few exceptions (e.g., Everitt & Fisher, 1995) and despite its affinity with neural network models of belief, coherentist insights have not been acknowledged. Similarly, even though according to two of its critics, "reliabilism is the most widely discussed contemporary epistemological theory" (Conee & Feldman, 2000), it has had little or no impact on the critical thinking field. There are objections to both coherentism and reliability. In the next sections we will explore how they can be addressed by an integration of the two approaches.

Complications

Despite its attractive features, externalism is far from trouble free. One of the problems cited by critics is spurious, as noted earlier: the alleged failure of externalism to provide practical guidance for cognizers. Among the other difficulties are some that pertain to the relationship between first-person and third-person points of view:

- Externalism fails to take account of residual internalist intuitions about the importance of what the cognizer *thinks*.
- What is the relationship between (i) estimates of confidence in beliefs produced *internally* by the relevant process and (ii) reliability estimates for the process as a whole produced by an external evaluator?

Other difficulties are inherent in the externalist paradigm itself:

- Any particular case of reasoning is an instance of more than one process. Thus, there is ambiguity about reliability estimates. Which process is the right one to use in evaluating the belief?
- There is a possibility that processes themselves are reliable by accident.

The most promising solution to all of these problems lies in the idea that externalist evaluation itself represents a *point of view*, as opposed to abstract objectivity. This insight must be combined with a careful study of how various points of view are coordinated with one another in a critical thinking dialogue.

The first and second problems concerns internalist intuitions and fairness in evaluation. Recall MAJ Jones, who has a highly reliable faculty for recognizing different types of tanks as a result of long training and experience. Suppose MAJ Jones does not realize how reliable her judgment is and indeed incorrectly believes it to be unreliable. From the internalist perspective, MAJ Jones is unjustified in accepting her own beliefs about tanks, even thought they are reliable (Bonjour, 1985). She would be wrong to accept them without very verification. Conversely, recall LT Smith. Her faculty for recognizing the presence of a T-62 is reliable only under a very narrowly defined set of conditions. But if LT Smith has no way of knowing what those conditions are or whether they obtain in a particular case, according to internalists her tank identifications are justified even though they are unreliable. LT Smith would be wrong not to accept her judgments! Both of these points have been taken to suggest that internalist intuitions based on fairness and duty, are not accounted for by externalism.

The second problem, concerning the relationships between internal and external judgments of reliability pertains to coherence. Judgments about reliability must be part of a network of beliefs that is evaluated with respect to its coherence. Thus, there is no escaping the kind of "circularity" emphasized by coherence theories (Sosa, 1991). Coherence theories stress the coherence of reliability judgments, while externalist theories stress the reliability of judgments based on coherence. But which is primary?

The third issue is the *generality* problem. The reliability of a cognitive faculty is simply its ratio of successes to failures under specified circumstances. But then, reliability depends on how generally or specifically the circumstances are specified (Conee & Feldman, 2000). If they are specified too generally, reliability is not very informative. For example, visually formed beliefs seem to be generally reliable; but visual pattern recognition processes that identify a nearby object as a tank in good conditions are much more reliable than the average visually formed belief. But should we also include the condition that dummy tanks exist in the area? If so, that same process is less reliable than the average visually formed belief. If we describe the actual present conditions with maximal specificity, reliability reduces to truth or falsity of the belief in the particular case. But justification should not entail absolute certainty; it should be possible to have a justified belief that is false or an unjustified belief that is true. How then is the appropriate level of generality chosen?

The fourth problem also pertains to the appropriate level of description of the process under evaluation. A cognitive process might be highly reliable, but if it was adopted by luck, e.g., without any insight into its effectiveness, the resulting judgments might seem less than fully justified. If the process resulted from a process that is itself reliable in producing reliable processes (e.g., learning or evolution), then justification is more complete. But how far back must reliability go in order for the products of a process to count as justified?

Points of View

Solution of these problems, and a reconciliation of reliabilism and coherentism, requires the recognition of two distinct points of view: the person whose knowledge is being assessed (call her the proponent P) and the person who is assessing that knowledge (call her the judge J). Judgments of reliability of P's beliefs are made by the assessor J. The assessor's purpose is quite straightforward. J would like to be able to use P's opinions as a source of information in a particular range of circumstances, but in order to do so must assess the extent to which P's beliefs can be trusted in those circumstances. J asks, for example: Can I infer from the fact that MAJ Jones believes this tank is a T-62 to the conclusion that it is a T-62? Can I infer from the fact that LT Smith believes there is a tank in the vicinity to the conclusion that there is a tank in the vicinity? J would like to infer from P's having a certain belief, that the belief is true and can be justifiably endorsed and adopted by J herself (Brandom, 2000, p. 120).

Distinguishing these two points of view enables us to resolve the coherence problem. From the point of view of the assessor J, judgments of the reliability of P must be arrived at just as other judgments are, by reference to their coherence with J's other beliefs and their fit to J's perceptual experiences. As Brandom (2000) puts it, concern with reliability is *external* only "because assessments of reliability (and hence of knowledge) can turn on considerations external to the reasons possessed by the candidate knower [P] himself." But assessments of reliability are *not* external to the reasons possessed by the assessor J. They inevitably occur within J's own system of beliefs, and coherence with those beliefs is a major determinant of J's conclusions

regarding the reliability of P. In dual-perspective reliabilism, second-order beliefs about reliability are required in order to anchor a coherent system of beliefs in reality. But it rejects the requirement that those second-order beliefs be part of the same system that is being evaluated.

Similarly, the generality problem arises only when reliability assessments are thought of as lacking a point of view, hence, as independent of both reasons and purposes. Since reliability is assessed from J's perspective, the scope of reliability assessments will depend on J's beliefs and purposes. In particular, reliability assessments will depend on (a) what J knows about the situation, (b) what J knows about P, and (c) the range of situations in which J might want to trust P as a source of information. If J is concerned with the trustworthiness of MAJ Jones' perceptual recognition of a tank and is aware of the presence of dummy tanks in the area, J will not regard MAJ Jones' judgment as reliable evidence for the presence of a tank. But if J trusts MAJ Jones generally, if the situations where dummy tanks are present constitute a small minority, and if J is not aware of the presence of dummy tanks in the area, then J will justifiably conclude that MAJ Jones' tank report is reliable. The issue is in part one of temporal scope and frequency of monitoring by, as illustrated in Figure 6.

The fairness problem is a matter of divergent purposes between internalist and externalist points of view. According to internalism, the purpose of critical thinking is to fulfill an intellectual duty, to carry out one's intellectual responsibilities in a blameless way. Thus, it is unfair to blame a critical thinker for disregarding relevant evidence if that information was not cognitively accessible (It is also unfair to credit her for ignoring evidence that was cognitively accessible, just because that information turned out to be inaccurate). But externalism shifts the purpose of critical thinking: It emphasizes the bottom line: accepting significant true beliefs and rejecting significant false ones. Because of this shift, there is no longer an issue of "fairness" in allocating praise and blame. Nonetheless, internalist intuitions about fairness can be captured in an externalist account by considering point of view. The candidate knower may assess the reliability of her own beliefs, adopting the perspectives both of judge J and subject of assessment P. Intuitions about fairness tend to correspond to the point of view of the candidate knower when evaluating the reliability of her own judgments. Both LT Smith and MAJ Jones made reasonable decisions based on the reliability assessments they made about their own judgments. J reached different conclusions simply because J used more information than they did.

But if the two perspectives can be combined within the same person, how can they remain distinct? Wouldn't reliability judgments be identical to the judgments arrived at by the first-order process? In other words, if a reasoning process inferred a probability of .8 confidence in a conclusion, wouldn't the assessment of the reliability of that belief also have to be .8, if it is done by the same person? The answer is no. The reality of the different viewpoints is confirmed in an experimental study by Leddo et al. (1990), in which different points of view were induced by assigning different roles to participants. Participants were asked to estimate the chance of success of a battle plan. Participants could be assigned the role of planners or of implementers. When participants performed as planners, they adopted an internalist point of view. They tended to estimate the chance of success by considering the possible *reasons* the plan might fail. This exercise helped them anticipate and plan for potential problems. But since the planners inevitably overlooked some possibilities, they overestimated overall chance of success. When participants performed as implementers, on the other hand, they adopted an externalist point of view. They tended to estimate chance of success statistically, by reference to the past frequency of success in

plans of a similar kind, not by trying to exhaustively enumerate failure scenarios. As a result, implementers were less overconfident.

The two points of view are distinct even when they are both embodied in the same individual. Critical thinking occurs *internally* by challenging a thesis or plan and making adjustments in response to problems that are found. In the internalist sense, critical thinking is an intrinsic part of reasoning. But critical thinking occurs *externally* by stepping back and questioning the reliability of the process as a whole under relevant conditions, in order to select the appropriate process, regulate its use of resources, and determine when confidence in the conclusion is high enough to stop. Since this kind of evaluation is done "from the outside," the process being evaluated may, but need not, involve reasoning; instead it might concern the accuracy of a perception, recall, or recognition. The two viewpoints draw on different kinds of information and involve different attitudes. They correspond to distinct but equally important levels of critical thinking.

Critical thinking research and teaching has paid scant attention to non-foundationalist viewpoints (Freeman, 2000). This is the reason that the concept of *argument* (with individual beliefs as conclusions) has occupied center stage. Non-foundationalist approaches such as coherentism and reliabilism shift the emphasis away from deliberative arguments about individual beliefs. Coherentism accounts well for the mutual adjustment of beliefs to one another in networks, but not for the special role of perceptual inputs or for computational limitations. Reliabilism accounts for beliefs in terms of the specific cognitive faculties that generate or sustain them, including both perceptual and inferential systems as they operate in real environments.

Reliability in Critical Thinking

The three-part model of critical thinking (Figure 9) integrates insights from coherentist and reliabilist theories of justification. The version of reliabilism depicted in Figure 31 has a foundationalist flavor because reasoning builds on a distinct, privileged class of beliefs generated by perception. By contrast, Figure 31 is a reliabilist framework that incorporates both coherentism and critical thinking. No beliefs are immune to revision based on incoherence with other beliefs. Perceptual systems produce *experiences* rather than beliefs, and these experiences are causal inputs to belief generating faculties. In other words, Figure 31 rejects the foundationalist assumption that there is a privileged class of beliefs that is immune to reasoning. On the other hand, it acknowledges that perceptual experience is an essential input to a coherence-based belief system (c.f., Haack, 1993; Thagard, 2000). The role of beliefs that are closely related to perceptual experiences is explained by appeal to their reliability, but it is not necessary for the *candidate knower* herself to have reflective second-order beliefs about her first-order beliefs, as coherence theories often require.

The three-part model of critical thinking forms the top tier of Figure 31, consisting of critical dialogue about mental models to achieve purposes under specific environmental conditions. Although critical thinking is reflective, it interacts with the more automatic operation of the coherence system. It takes sets of beliefs from the coherence system as inputs, creates and critically evaluates mental models, and in turn feeds its conclusions back as inputs to the coherence system. All cognitive faculties – perception, coherence-based reasoning, and critical thinking – are designed to reliably achieve particular purposes in particular environments in consort with each other. Judgments of reliability may be made from an external point of view, to

determine whether another person's opinions can be trusted, or may be made internally (but still, from a hypothetical "outside" point of view) to regulate use of one's own faculties in knowledge acquisition.

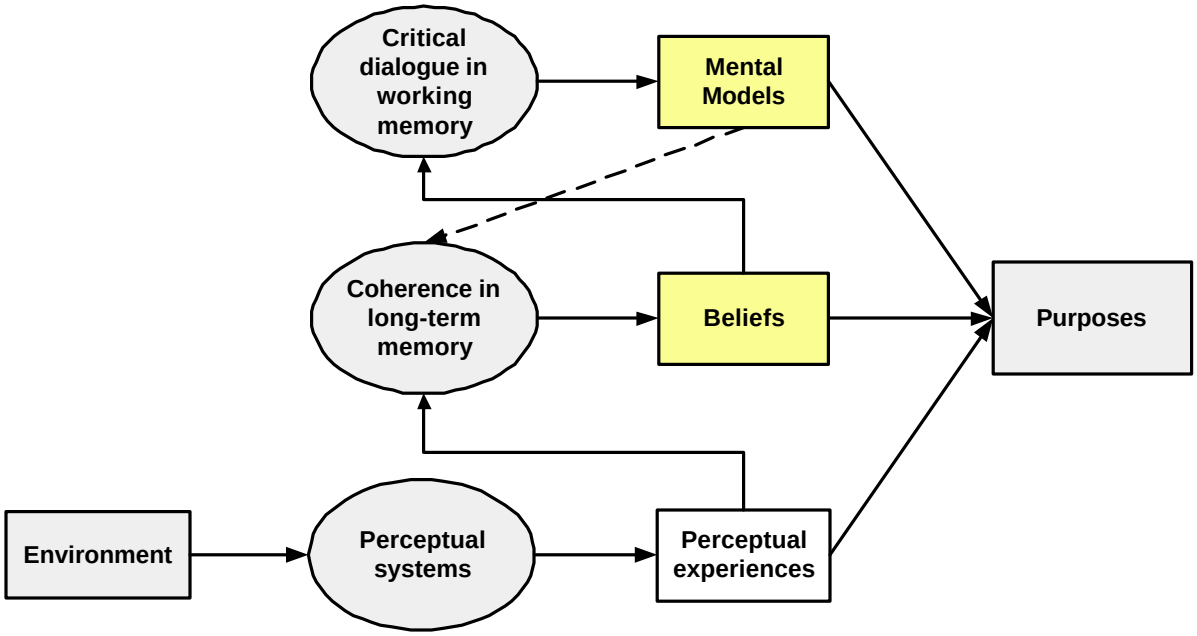


Figure 31. A reliabilist framework that integrates a foundationalist theory of perceptual evidence, a coherence theory of inference, and a critical thinking model of reflective reasoning.

11. EVALUATION OF DIALOGUES IN TEAMS

Communication and Shared Mental Models

Identifying and improving team processes has been the focus of ongoing attention from researchers. An advantage of the use of teams is that in many situations they produce better problem-solving and decision-making outcomes than do individuals working alone. The improved outcomes are partially due to team interaction gains that result from pooled individual information. One of the consistent conclusions has been that communication is an essential aspect of team decision making (Klein, 1998; Duffy, 1993; Orasanu & Salas, 1993; Means, Crandall, Salas, & Jacobs, 1993). For example, Orasanu and Salas (1993) pointed how the importance of explicit communication in decision making situations: "...good crews were much more explicit in defining the problem, articulating plans and strategies for coping with it, obtaining relevant information explaining the rationale, and allocating and coordinating responsibilities among the crew." (Orasanu & Salas, p. 334). However, communication per se is also one of the least explicated constructs and the least well modeled in simulations.

During training, experts must communicate to the team members not only their assessment of the situation, but the assumptions and experiences that led them to that assessment. By exposing novices and other team members to their mental model, leaders provide the opportunity for the others to incorporate it into their own mental models, creating situational mental models that include terminology, assumptions, and thinking patterns. According to Orasanu and Salas (1993), groups were more successful in solving problems if they had analyzed problems instead of focusing on solutions. Intensive practice would serve to fix those responses. Later, in a critical situation, team members would be able to recall those responses and apply them accordingly.

In addition to acquiring domain knowledge in training, the team could also be inculcated in team norms that would be operating during a critical incident situation. For example, through feedback and rewards, trainers could reward lower-status team members who showed persistence in questioning higher-status team members, to overcome a danger that Orasanu and Salas (1993) identified in critical-incident decision-making situations.

Shared mental models will be more valuable if the models are created prior to teams being exposed to high-pressure situations (Figure 32). Problem solving will suffer in teams whose members have low tenure, have varied levels of domain knowledge, are unfamiliar with their leader's abilities, have low organizational or goal commitment, etc. For such a team, the structure imposed by the process would be essential in developing a shared mental model, not only of situational responses, but more importantly, of the communication patterns inherent in the team.

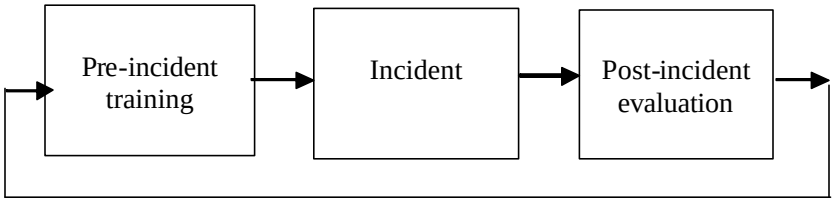


Figure 32. Points for enhancing shared mental models.

Domain knowledge, which refers to the declarative and procedural knowledge of team leaders and team members, is essential to effective decision making. But in addition, team decisions are made in a social context; thus, team activities have to be coordinated for the team to be effective. Effective team coordination in turn is enhanced when team members share mental models (Paris, Salas, & Cannon-Bowers, 2000). Communication among team members, through validation and elaboration of shared mental models, leads to improved decision making and problem solving. As such, it is a crucial component of any simulation of human decision making.

According to Senge (1990, p. 8), "Mental models are deeply ingrained assumptions, generalizations, or even pictures or images that influence how we understand the world and how we take action". Green (1996, p. 120) said that decision making is a social process requiring participants to ". . . be able to represent the various arguments, keep track of them, and be able to contribute to them . . . Mental models need to represent social situations and the way in which actions are achieved through talk."

Anderson, Howe, and Tolmie (1996) suggest that transitory mental models are constructed and negotiated during problem-solving dialogue. The researchers found evidence that individuals differ in their ability both to build, maintain in memory, and test individual transitory models, as well as to negotiate joint transitory models (Anderson et al., p. 270). Their study (1996, p. 268) found that shared common frames of reference, including shared vocabulary, facilitated the process of negotiating a shared model. The authors concluded that the participants engaged in social construction and internalization of a joint model during conversation. The participants also showed evidence of incorporating the work group experience into their personal, enduring mental model of that domain. Anderson et al. also suggested that the mental models that result from the interaction between individuals are affected by the social and culture contexts within which they operate.

The Relevance of Dialogue Theory

One approach to modeling team communication and related problem-solving activities that has never been explicitly examined before is that of incorporating the principles of dialogue theory. Informal logic more generally analyzes methods that are used to interpret and evaluate arguments. Dialogue theory focuses more explicitly on the communicative contexts in which such arguments are embedded, including the team interactions by means of which mental models are created, validated, and shared. van Eemeren et al. (1996, p. 193) contend that, "...argumentation is the practice of justifying decisions *under conditions of uncertainty*".

Furthermore, "... argumentation should be put in the social context of a process of joint problem solving." (p. 277).Social arguments are how social groups identify disagreements, negotiate agreements, or reach consensus based on a background encompassing shared values, meanings, and problems. These arguments are dialectical because they are interactive, involving two-way exchange, building upon previous exchanges. Because the meanings of the arguments are a function of their purposive contexts, the arguments are pragmatic. The arguments are supported by contextual rules and understandings and are directed by intentions toward a goal (van Eemeren, Grootendorst, & Henkemans, 1996, p. 164-165).

Particularly appropriate to team application is a form of persuasion dialogue, called critical discussion, which specifically considers how to resolve a difference of opinion between participants. In the team problem solving and decision making process, dialogue theory may improve our ability to model communication skills, by representing elements (that may or may not be present in an actual team) of a shared mental model of the critical discussion process. Dialogue theory helps capture the give-and-take that enables teams to make collective decisions, generate transitory shared models of the situation, and increase the extent of their overall shared knowledge.

Individual performance can be improved when the structure of the process has been incorporated into mental models and into the repertoire of critical thinking skills. An individual who can both formulate sustainable premises and recognize their presence in the arguments of others will be able to participate more effectively in team decision-making activities. In typical teams, members who have a formal system for questioning the critical reasoning of others, even in the face of non-conflict or hierarchical norms, are able to employ those skills where others are reluctant to do so. Additionally, in situations where highly trained teams are engaged in mission-critical activities, e.g., command centers, the possession of highly internalized critical discussion skills enables them to respond more quickly and effectively to decision-making requirements of situations encountered.

Although not completely synonymous, the terms critical thinking and critical discussion are used interchangeably. In the context of the article, the term critical discussion will refer both to internal critical thinking processes and to external dialogues.

Team Decision Making

According to Guzzo (1995, p. 4), team decision-making is "... interconnected activities that include gathering, interpreting, and exchanging information; creating and identifying alternative courses of action; choosing among alternatives by integrating the often-differing perspectives and opinions of team members; and implementing a choice and monitoring its consequences." Team decision making has four components: situation assessment, metacognition, shared mental models, and resource management (Orasanu & Salas, 1993).

According to Walton (1998, p. 34), "the goal of the dialogue is defined by its originating issue, which is the problem, question, or conflict the dialogue is supposed to solve, answer, or resolve." Ilgen, Major, Hollenbeck, and Sego (1995, p. 138) point out that decision making is a subset of problem solving,

> "to solve a problem, people must generate issues or dimensions on which they will seek information: then they must reach a decision regarding their approach to the problems. For decision making, the information set is fixed, and individuals

make decisions or choices based on a fixed set. Although it is readily accepted that decision making is a subset of problem solving, and that many more problems are likely to be of the problem-solving than of the purely decision-making nature there tends to be little integration among the works in problem solving with those in decision making."

But the decision-making process is constrained by resource limitations. According to March and Simon (1958, as cited in Ilgen, Major, Hollenbeck, & Sego, 1995), people "satisfice" rather than "optimize" when making decisions. That is, they search for a few alternatives that meet the specified criteria, instead of trying to identify all possible alternatives. Tjosvold (1995, p. 89), stresses the importance of constructive controversy to decision making: "Contrary to the common assumption that cooperative goals promote harmony and avoidance of conflict, the theory of constructive controversy proposes that open discussion of opposing views is most critical for making cooperative situations productive..." Therefore, teams that follow the process of critical discussion during decision making and problem solving will presumably improve the quality of their output.

Some of the various factors presumed to influence team decision making are listed below: (Ilgen, Major, Hollenbeck, & Sego, 1995; Klimoski & Jones, 1995).

- Individual member skills: communication skills, task vs. interpersonal orientation, experience with similar situations, expertise, investment in (commitment to) outcomes, age, tenure, preference to make decisions quickly on limited information, social-interaction skills, knowledge, motivation, role expectations.
- Team dynamics: clarity of communication, member roles, member hierarchy, shared mental models (understanding), team composition, established communication process, e.g., one sentence at a time or developed argument, trust, cooperation, coordination, differential value of individual member input to decisions, groups norms, group size, group composition, interpersonal dynamics, homogeneity of attitudes, values, and preferences.
- Organizational specifics: staffing practices, reward systems, organizational climate, intergroup interactions (Guzzo, 1995).
- Situation specifics: time available, complexity of decision, clarity of purpose/goal.

The following table gives a breakdown by category of some of the characteristics that are presumed to influence the quality of the decision-making process and hence the outcomes of the process (Cohen, Freeman, & Thompson, 1998; Ilgen, Major, Hollenbeck, & Sego, 1995; Klimoski & Jones, 1995; Klein, 1998). The decisions and outcomes will be of higher quality the greater the degree to which the characteristics are representative of the individuals, leader, group, and organization. However, in the case of the situation, the greater the degree of the characteristic, the more difficult it will be to identify an optimal or favorable decision and outcomes.

Level	Characteristic
Individual	
	Cognitive ability
	Confidence level in abilities and knowledge

	Communication skills, e.g., ability to express self clearly
	Commitment to group goals
	Commitment to organization goals
	Decision-making style
	Degree of autonomy and initiative and willingness to express both
	Domain knowledge
	Experience with similar situation
	Experiential (vs. abstract) type of domain knowledge
	Level of assertiveness
	Long-term time orientation
	Systems viewpoint orientation
	Problem-solving skills
Leader	
	Assertiveness level
	Communication skills, e.g., ability to express self clearly
	Confidence level in own/team abilities
	Commitment to group goals, i.e., ownership of outcome
	Commitment to organization goals
	Domain knowledge type, e.g., abstract or experiential
	Domain knowledge: declarative, procedural
	Experience with similar situations
	Expected support from outside influences, e.g., superiors, constituencies
	Expertise
	Leadership style, e.g., authoritarian, participative, situational
	Leadership style: supportive feedback style
	Orientation: short term or long term
	Preference to not make decisions if time is available for deliberation
	Problem-solving skills
	Tolerance for ambiguity, e.g., willingness to decide on incomplete information
Team Dynamics	
	Cohesiveness of team
	Cooperative (vs. competitive) orientation (norms)
	Commitment to team goals
	Commitment to organizational goals
	Confidence in leader's abilities
	Degree of exposure to similar situations, i.e., experience with similar situations
	Homogeneity of team members, e.g., demographic variables, experience, values

	Norms of team toward participation, commitment level, etc.
	Overlap of understanding of other team member's responsibilities
	Participative (vs. hierarchical) communication pattern (norms)
	Participative (vs. hierarchical) problem-resolving pattern (norms)
	Perceived support by management/superiors
	Roles and responsibilities of team members clear
	Shared mental models
	Tenure as an intact team, e.g., frequency of turnover of team members
	Team training: in group processes, in procedures
	Trust between members
Organization	
	Degree to which superiors will support team outcomes
	Intergroup interactions
	Organizational climate
	Purpose for which the team was assembled, e.g., manufacturing, crisis resolution, which is reflected in centrality of decisions to be made
	Reward systems
	Staffing practices
	Supporting processes: effectiveness and efficiency
	Supporting technology: effectiveness and efficiency
Situation Constraints	
	Ambiguity as to exact nature of problem/situation
	Ambiguity as to goals
	Complexity of situation
	Complexity of decision variables
	Dynamism, i.e., changing nature, of situation
	Lack of information about situation
	Lack of time available to make decision
	Risk associated with decision
	Severity of outcome consequences

Process of Critical Discussion

There are various types of argumentative dialogue, each with different goals (Walton, 1998, p. 31). The types of dialogue are persuasion dialogue (a.k.a. critical discussion), information seeking (interview, advice-solicitation, expert consultation), negotiation, inquiry (scientific, public), eristic (quarrel), and deliberation. As noted earlier, the critical discussion is especially appropriate to the team decision-making process because its goal is the mutually acceptable resolution of differences of opinion (van Eemeren, Grootendorst, & Henkemans, 1996, p. 278).

A critical discussion involves the rules applicable in each of the four resolution stages, in which premises and conclusions are subjected to critical analysis. The four stages of resolving a disagreement correspond to the four phases of an argumentative exchange, as outlined below. The four stages (Figure 33) can apply to one standpoint or to a series of sequential standpoints. (van Eemeren, Grootendorst, & Henkemans, 1996, p. 280).

- Confrontation stage: A standpoint is expressed and opposition is raised to it. (When a standpoint is contradicted or doubted, disagreement results.)
- Opening stage: protagonist defends his/her standpoint, while the antagonist critiques the standpoint and its defense. (A critical discussion cannot occur unless the parties can exchange views, which requires a common ground of shared knowledge, rules, values.)
- Argumentation stage: antagonist challenges defenses raised by the protagonist. (As long as one party is doubtful of the other's argument, he or she continues to critically appraise the other's arguments.)
- Concluding stage: either the protagonist or antagonist concedes inability to support the standpoint and withdraws it. (If both parties do not agree on the outcome, then the difference of opinion was not resolved.)

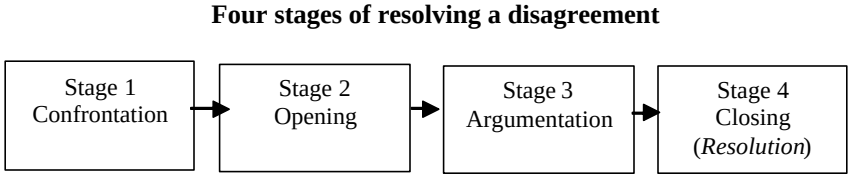


Figure 33. Stages of critical discussion.

Rules of Evaluation and Errors

For each of the four stages, there are rules of evaluation that must be followed. See the Normative Evaluation section below and the tables at the end of this chapter for an enumeration of the rules and errors according to van Eemeren et al. and Walton.

The rules are norms for critical discussion. Errors impede the resolution of the difference of opinion. For example, a protagonist must be prepared to rationally defend his or her standpoint, and a position that cannot be rationally defended must be withdrawn (Walton, 1998). An antagonist or protagonist can only make one type of response, e.g., single question, at a time, and the other must respond appropriately with a matching response. Protagonists and antagonists may assert, concede, ask, or retract during the dialogue.

For example, at any point in the discussion, members may not be participating completely, e.g., deliberately withholding information, failing to challenge dubious claims, overtly agreeing but covertly disagreeing, concealing agenda, etc. (Walton). There are additional

errors that can occur due to conditions in the situation, or at the organization, team, or individual level. At the individual level, team members may not understand the process well enough to participate, may lack the cognitive skills necessary to participate, may lack training in problem-solving or decision-making skills that would enable them to understand the conceptual framework of the discussion. At the team level, conflicting demands made upon the members may inhibit their full attention and participation. At the situation level, distracting environmental conditions may result in the discarding of trained procedures. Any such deficiency will result in degradation of the shared mental model, at least in the sense that the model will not be shared to the extent possible or desirable.

Walton (1998, pp. 249-252) proposed a four-step method for evaluating arguments. The model, expanded to include details from the critical thinking model proposed by Jones (1996), is presented below:

1. Identify and evaluate the premises and conclusions of the argument. (Understand the structure and analysis of arguments, differentiate between fact and opinion, examine assumptions.)
2. Identify the contexts of the dialogue, e.g., goal, type (persuasion, negotiation, information seeking, etc.).
3. Assess the burden of proof.
4. Recognize and evaluate informal fallacies (Avoid incorrect reasoning, fallacious arguments, ambiguity, and manipulative reasoning.)
5. Be flexible and open minded when looking for explanations, causes, and solutions to problems.
6. Focus on the whole picture, while examining the specifics.

Note that each step has multiple sub points.

A Model

Can the rules for critical discussion under persuasion dialogue be used to describe issue resolution in the team decision-making process?

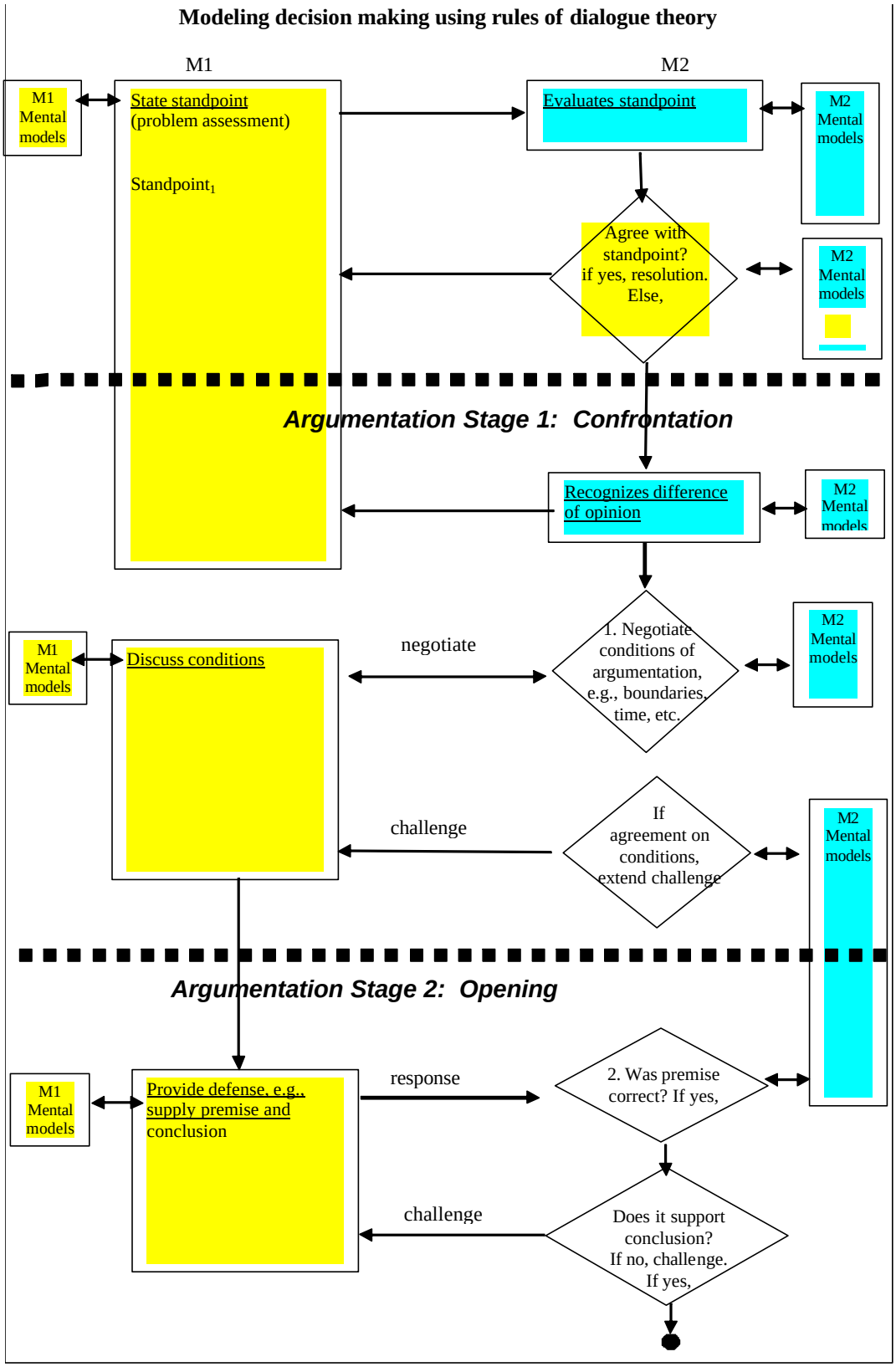
Assumptions

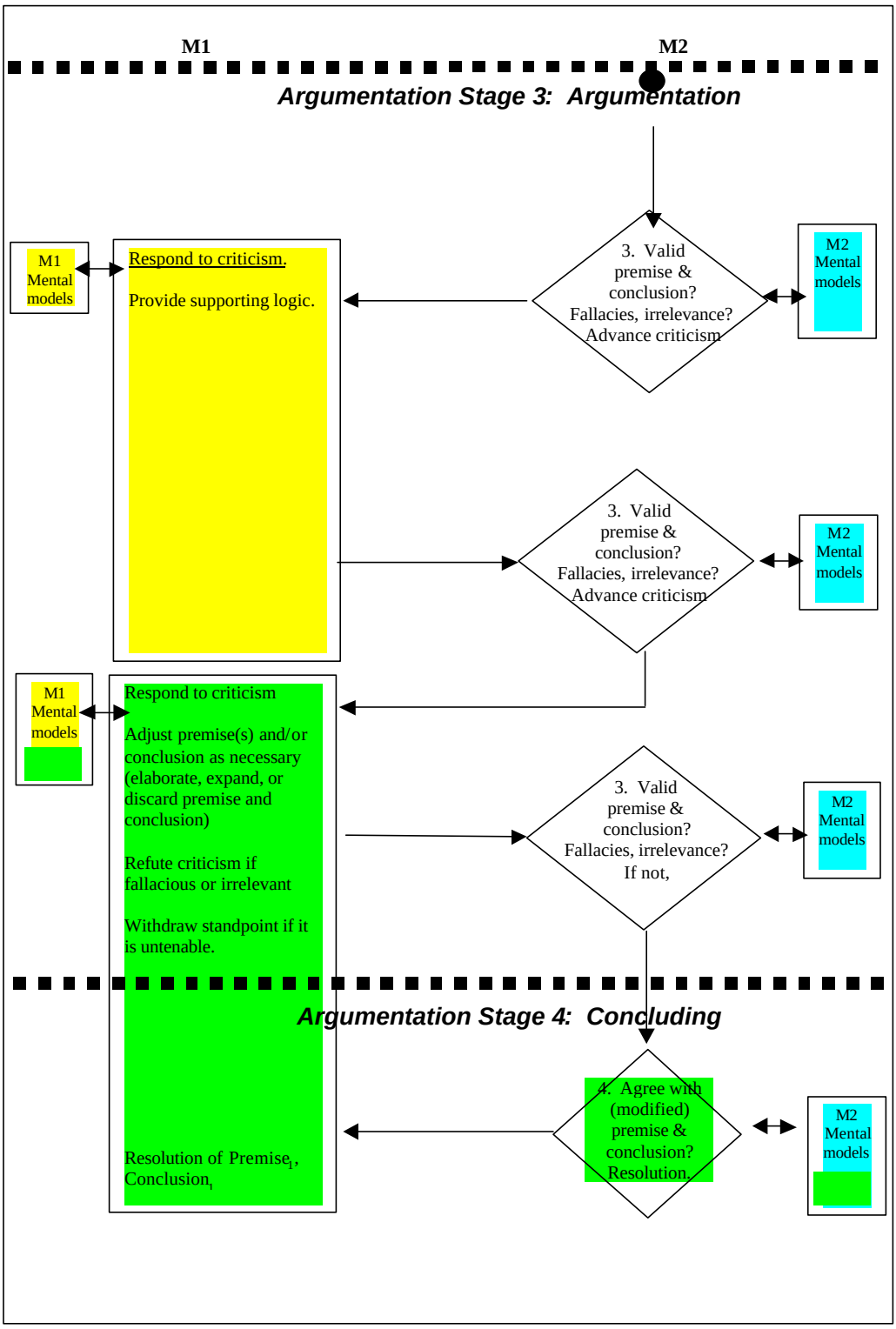
- In the (very) simplified model shown below, there are two team members: M1: Team Member 1 is the protagonist, M2: Team Member 2 is the antagonist.
- The burden is on Team Member 1 to convince Team Member 2 of his/her premises and conclusions.
- The goal is to determine whether the protagonist's standpoint can be maintained in the face of criticism.
- Team Member 2 can agree with, disagree with, or be neutral to any proposition put forth by Team Member 1.
- The model indicates the four stages of resolving an argument.

- The process is sequential, and numerous sub-arguments may be resolved within a larger argument.
- The decision-making components explained by the model are
    a. situation assessment between Team Member 1 and Team Member 2,
    b. metacognition processes evinced by Team Member 2's responses to Team Member's premises and conclusions, and
    c. changes in shared mental models as resolution to disagreement is reached.
- Members are assumed to have knowledge of and facility with interpersonal communication such that they will advance a challenge in a non-threatening manner and recognize cultural perspectives, e.g., "face saving".

Constraints

- Failure by team members to follow the "rules" of critical thinking or limitations of individual team members in recognizing ambiguity, lack of clarity, fallacious premises or conclusions will lead to the acceptance of invalid arguments. Decisions based on invalid arguments will result in lower-quality output by the team. Limitations of individual team members may be compensated for by strengths of other members. *Ideally, team members would be highly trained in the process of critical discussion.*
- The model shown below is consistent with the meta-recognitional cycle for reducing uncertainty that was proposed by Cohen, Freeman, and Thompson (1998), which tests first for incompleteness, then tests for conflict, and then tests for unreliability.


Modeling decision making using rules of dialogue theory
M1
M2
M1 Mental models
State standpoint (problem assessment)
Standpoint1
Evaluates standpoint
M2 Mental models
Agree with standpoint? if yes, resolution. Else,
M2 Mental models
Argumentation Stage 1: Confrontation
Recognizes difference of opinion
M2 Mental models
M1 Mental models
Discuss conditions
negotiate
1. Negotiate conditions of argumentation, e.g., boundaries, time, etc.
M2 Mental models
challenge
If agreement on conditions, extend challenge
M2 Mental models
Argumentation Stage 2: Opening
M1 Mental models
Provide defense, e.g., supply premise and conclusion
response
2. Was premise correct? If yes,
challenge
Does it support conclusion? If no, challenge. If yes,



Process and Outcome Evaluation

Critical discussion processes can be evaluated normatively, keeping in mind that the purpose of critical discussion is an exchange of speech acts to resolve a difference of opinion.

Normative Measurement

The following rules must be observed during a critical discussion for it to be a legitimate discussion. The rules indicate how much the actual critical discussion deviated from the ideal discussion that would best lead to resolution of the difference of opinion (van Eemeren et al., 1996). Note that inductive and deductive reasoning skills, and underlying cognitive processes are not specified.

The method of evaluation would be to analyze a critical discussion and each sub-argument of a critical discussion, using a checklist based on the following list of rules. Any deviations would indicate that the process was not followed.

General rules.

- The participants must ignore superfluous and immaterial comments, e.g., elaboration, immaterial interruptions, sidelines, and unnecessary repetitions.
- Equivocation and ambiguity must be rejected.
- Shifts between dialogue types are not permitted.
- The participants must use transitional phrases to facilitate listener comprehension, e.g., indicate shift from premise to conclusion, by use of clauses such as "because..., therefore..."
- Each participant must demonstrate critical thinking discussion skills: recognizing disagreement, voicing disagreement, following the rules, recognizing if rules are being broken, voicing broken rules, responding appropriately according to stage of argumentation, avoiding overlap between stages, providing and interpreting verbal cues correctly, etc.

Argumentation characteristics.

- The type of discourse, e.g., persuasive argumentation, must be determined so that the appropriate analysis can be made.
- The nature of the disagreement, e.g., single non-mixed, must be identified.
- The protagonist and antagonist must be identified.
- The standpoint at issue must be identified.
- The argument must be identified so that the premises and conclusions may be analyzed.
- The argument structure unifying the argumentation must be identified so that the protagonist's overall defense can be evaluated.
- The premises that support the standpoint must be identified so that the argumentation scheme, e.g., similarity, that links the two can be evaluated.

Steps of the four-stage process.

1. Confrontation stage

- The protagonist must express a standpoint.
- The antagonist must recognize a conflict, and must indicate disagreement by voicing a difference of opinion, i.e., challenging the standpoint.

- The antagonist must ask for amplification, precision, definition, etc., to make explicit those elements that were implicit.

2. Opening stage

- The protagonist and antagonist must agree on the discussion rules.
- The protagonist must at every stage either uphold or modify the standpoint.
- The antagonist must ask for amplification, precision, definition, etc., to make explicit those elements that were implicit.

3. Argumentation stage

- The protagonist must advance argumentation, e.g., premises, after determining if the challenge was valid.
- The antagonist must identify the points under question and ask for additional argumentation if the reasoning is invalid.
- The antagonist must accept or reject the protagonist's argumentation.
- The antagonist must ask for amplification, precision, definition, and to make explicit those premises that were implicit.
- The premises must support the main goal of the discussion.

4. Concluding stage

- The protagonist must uphold or retract his or her statement.
- The antagonist and the protagonist must agree on the result of the argument.
- If the antagonist does not agree with the defense put forth of the protagonist's standpoint, the standpoint must be withdrawn.
- The antagonist must ask for amplification, precision, definition, etc., to make explicit those elements that were implicit.

Evaluative Measurement

The use of the critical discussion process by a team during a decision-making process can be evaluated both at an overall level and at an elemental level.

There are several categories of evaluation: overall; process evaluation, including mechanics such as overall questioning strategy during the four-stages; skills evaluation; outcome evaluation; participant self-evaluation.

Overall evaluation of entire discussion as element of problem-solving model.

Yes No	Describe	Evaluation Item
		At each stage of the problem-solving model: define problem, identify alternatives, select solution, was the model employed appropriately?
		Did the critical discussion model support the problem-solving model?

Overall evaluation of entire critical discussion.

Yes No	Describe	Evaluation Item
		How many sub-arguments occurred?
		Did each contribute to the overall solution?
		Did the sequence of reasoning support the conclusion?
		What general strategy was followed?
		What pattern of question-and-response followed?

Critical discussion incident evaluation.

For each specific incident of critical discussion:

Yes No	Describe	Evaluation Item
		Was the type of dialogue correctly identified? What was it?
		What was the exchange type, e.g., single non-mixed?
		What was the structure of argumentation?
		What was the scheme of argumentation?
		Was the protagonist correctly identified? Who was it?
		Was the antagonist correctly identified? Who was it?
		How often did the dialogue switch from one type to another? How often were the switches between dialogue types justified?

Stage 1: Confrontation

Yes No	Describe	Evaluation Item
		What was the standpoint at issue?
		Was standpoint articulated clearly? Did it include any premises? Were they clear?
		What was protagonist's starting point? Did it support the conclusion?
		Did the antagonist recognize a difference of opinion? Was it expressed according to the rules? If not, did the protagonist identify the error?
		Did the participants make superfluous and immaterial comments? Elaborations? Immaterial interruptions? Sidelines? Unnecessary repetitions? Did the other respond appropriately?
		Did a shift between dialogue types occur?
		Were the rules of the step followed? To what degree?

Stage 2: Opening

Yes No	Describe	Evaluation Item
		Did the protagonist respond appropriately to the challenge?
		Were premises explicitly stated or did the antagonist have to ask for clarification?
		What was the protagonist's response? Was it appropriate? Did it further the discussion? If an error was committed, what was it and how did the antagonist respond?
		Did the participants make superfluous and immaterial comments? Elaborations? Immaterial interruptions? Sidelines? Unnecessary repetitions? Did the other respond appropriately?
		Did either party use equivocations?
		Did either party use ambiguity?
		Did either party attempt to shift the dialogue type? Who?
		Were the rules of the step followed? To what degree?

Stage 3: Argumentation

Yes No	Describe	Evaluation Item
		What was premise 1? What was challenge 1? What was defense 2?
		What was premise 2? What was challenge 2? What was defense 2?
		How well supported were the premises? Were premises sound?
		How well supported were the conclusions?
		How many irrelevancies appeared during the discussion? What were they?
		How many fallacies were identified and resolved? What were they?
		How many ambiguous language issues were resolved? What were they?
		Did either practice equivocation?
		Did the participants make superfluous and immaterial comments? Elaborations? Immaterial interruptions? Sidelines? Unnecessary repetitions? Did the other respond appropriately?
		Were the rules of the step followed? To what degree?
		Did shifts between dialogue types occur? What were they?

Stage 4: Concluding

Yes No	Describe	Evaluation Item
		Was conclusion reached, i.e., agreed to by protagonist and antagonist?
		Did the protagonist withdraw his/her standpoint if it was unsuccessfully supported?
		Were the rules of the step followed? To what degree?
		Did the participants make superfluous and immaterial comments? Elaborations? Immaterial interruptions? Sidelines? Unnecessary repetitions? Did the other respond appropriately?
		Was there a shift between dialogue types?

<u>Skills evaluation.</u>

Low Med High	Comments	Evaluation Item
		For each participant, to what degree were the skills underlying successful use of the critical thinking process used?
		recognized disagreement
		followed the rules,
		voiced disagreement
		recognized if rules were being broken
		voiced broken rules
		responded appropriately according to stage of argumentation
		avoided equivocation
		avoided ambiguity
		avoided overlap between stages
		Used transitional phrases to facilitate listener comprehension, e.g., indicate shift from premise to conclusion, by use of clauses such as "because..., therefore..." provided verbal cues correctly
		interpreted verbal cues correctly
		Use inductive logic correctly?
		Use deductive logic correctly?

Outcome evaluation.

Yes No	Describe	Evaluation Item
		Mental models – Is there evidence that mental models changed as a result of the discussion?
		Does agreement with the conclusion indicate that both participants share at least some aspects of the temporary mental model?

Participant evaluation.

Low Med High	Comments	Evaluation Item
		SME or observer evaluation:
		• Team's performance
		Overall
		Underlying skills
		Process rules
		Process steps
		Strategy
		• Leader's performance
		Overall
		Underlying skills
		Process rules
		Process steps
		Strategy
		• Individual team member's performance
		Overall
		Underlying skills
		Process rules

		Process steps
		Strategy
		Leader's evaluation
		Of team
		Of outcome
		Of own performance
		Of each team member
		Team's evaluation
		Of team
		Of outcome
		Individual evaluation
		Of own performance

Additional Research Questions:

1. After how much practice does critical thinking becomes an automatic process by team members,?
2. To what extent does the underlying level of subject-matter knowledge influence the quality of problem solving and decision making during argumentation?
3. Are there personal characteristics of team members that inhibit their use of critical thinking, e.g., tendency to see arguments as extremes, amount of cognitive ability, etc.?
4. Are there situations in which the critical thinking process is detrimental, e.g., time-dependent situations in which the most knowledgeable person is best qualified to make a decision?
5. To what extent does the process of argumentation improve the subject-matter mental models of participants?
6. Does the use of technology to solve problems and make decisions, e.g., emails or phone calls instead of face-to-face conversation, interfere with the use of the critical thinking process?
7. Does the use of the critical thinking process result in more alternatives? If so, are the increased alternatives of better quality?
8. Does the use of critical thinking transport from one team setting to another? What problems are encountered if a team member trained in critical thinking attempts to use the critical thinking process in a team that has not been trained in its use?
9. Does individual team member satisfaction increase with the use of the critical thinking process?
10. To what degree do strong communication skills, e.g., clarifying assumptions when speaking, automatically defining terms that may not be understood by team members, monitoring responses for behavior consistent with communications, etc., underlie the decision-making process?
11. Is the critical thinking process incompatible with the creative thinking process? Can critical thinking be suspended during creative thinking sessions?
12. How can the quality of communication be evaluated?

13. How do team members evaluate the competency of their team mates? How do and to what degree do leaders identify and rate weaknesses in team members, e.g., declarative knowledge, procedural knowledge, communication skills, etc.? How thoroughly do leaders incorporate their perceptions of team member strengths and weaknesses when making decisions?
14. How can the equivalence of shared mental models be evaluated?

According to van Eemeren et al.

The information in the following table is from van Eemeren et al. (1996, Chapter 10). Each of the ten rules is a norm for critical discussion. Any violation of the rules threatens the resolution of the difference of opinion. van Eemeren et al. define fallacy as a speech act which frustrates efforts to resolve differences of opinion. The type of discussion in the table assumes a single non-mixed argumentation structure, i.e., the protagonist must defend his or her standpoint.

Actions	Rules	Obligatory Speech Acts	Errors & Fallacies
1.1.1.1.1.1 General			
• Identify which of the six types of dialogue is being undertaken. Refer to the rules specific to that type of discussion. When violations of the rules occur, the rule violated must be identified so that the appropriate criteria for satisfying can be met.	• The speech act must agree with the critical discussion rules. • The roles of antagonist and protagonist must be clear. • The argument structure, e.g., single, non-mixed, must be clear.		Protagonist or antagonist: • Deliberately withholds information • Fails to challenge dubious claims • Overtly agrees but covertly disagrees • Conceals a personal agenda
1.1.1.1.1.2 Stage 1 - Confrontation			
The antagonist recognizes disagreement with the protagonist's standpoint in such a way that the nature of the difference of opinion is clear. Both parties must plausibly assume that a critical		• Expresses a standpoint • Accepts or fails to accept a standpoint • If fails to accept, upholds non-acceptance of standpoint • Defines, makes precise,	

discussion is occurring, the goal of which is resolution of the difference of opinion.		amplifies, etc. • Asks for elaboration.	
	• "Rule (1) Parties must not prevent each other from advancing standpoints or from casting doubt on standpoints. "		*Protagonist or antagonist* • Banns a standpoint • Declares a standpoint to be sacrosanct • Puts pressure on other by threatening with sanctions or by inducing feelings of compassion • Makes a personal attack on the other: a) depicts person as bad, stupid, unreliable, b) casts suspicion on motives of other, c) points out inconsistencies between party's current and /or past deeds and ideas. The result of the first error is to restrict the standpoints that can be criticized or advanced. The result of the second is pressure the opponent, appeal to the opponent's compassionate feelings, or to discredit the opponent (expertise, integrity, credibility, impartiality)
Misunderstandings can result from incorrect interpretations of vague,	• "Rule (10) A party must not use formulations that are		*Protagonist or antagonist* • Takes unjust advantage of

unclear, or ambiguous formulations. Misunderstandings can lead to a pseudo-resolution. Premises must, therefore, be expressed explicitly.	insufficiently clear or confusingly ambiguous and a party must interpret the other party's formulations as carefully and accurately as possible. "		unclearness: a) uses structural unclearness, b) uses implicitness, c) uses indefiniteness, d) uses unfamiliarity, e) uses vagueness • Takes unjust advantage of ambiguity: a) uses referential ambiguity, b) uses syntactic ambiguity, c) uses semantic ambiguity.
1.1.1.1.1.3 Stage 2 - Opening			
The roles of protagonist and antagonist are assigned.		• Accepts challenge to defend standpoint • Decides to start discussion • Agrees on discussion rules • Challenges to defend standpoint • Defines, makes precise, amplifies, etc. • Asks for elaboration.	
	• "Rule (2) A party that advances a standpoint is obliged to defend it if asked by the other party to do so. "		*Protagonist* • Evades the burden of proof: a) presents a standpoint as self-evident, or b) personally guarantees the correctness of the standpoint or c) inoculates the standpoint against criticism. • Shifts the burden of proof: a) tries to make the antagonist show

			that the standpoint is wrong. The result of the error is that the protagonist tries to give the impression that the standpoint shouldn't be questioned or doesn't need to be defended.
Misunderstandings can result from incorrect interpretations of vague, unclear, or ambiguous formulations. Misunderstandings can lead to a pseudo-resolution. Premises must, therefore, be expressed explicitly.	• "Rule (10) A party must not use formulations that are insufficiently clear or confusingly ambiguous and a party must interpret the other party's formulations as carefully and accurately as possible."		<u>Protagonist or antagonist</u> • Takes unjust advantage of unclearness: a) uses structural unclearness, b) uses implicitness, c) uses indefiniteness, d) uses unfamiliarity, e) uses vagueness • Takes unjust advantage of ambiguity: a) uses referential ambiguity, b) uses syntactic ambiguity, c) uses semantic ambiguity.
The notions of acceptability (providing a preferred response) and disagreement (providing opposition) are agreed upon, either formally or informally. The protagonist is obliged to defend his/her controversial standpoint. The antagonist is obliged to critically evaluate the protagonist's standpoint and its defense.			
The parties determine			

the existence of a common ground so that they can productively exchange viewpoints. Such a starting point includes shared background knowledge, values, etc.			
Keeping in mind the goal of resolving a difference of opinion based on its merits, the identity and correctness conditions of the speech acts appropriate for the type of dialogue are identified.			
1.1.1.1.1.4 Stage 3 – Argumentation			
The protagonist advances an argument to defend his or her standpoint, to which the antagonist, if he or she is not convinced of the merits of the argument, advances a critical appraisal. This cycle repeats itself.		• Advances argumentation • Accepts or fails to accept argumentation • Requests argumentation • Defines, makes precise, amplifies, etc. • Asks for elaboration	•
To resolve a difference of opinion, both parties need to be addressing the same standpoint. The central issue must not be distorted by either party. The parties must discuss the argumentation logically, not	• "Rule (3) A party's attack on a standpoint must relate to the standpoint that has indeed been advanced by the		*Protagonist* • Imputes a fictitious standpoint to the other party: a) advances the opposite standpoint as one's own, b) refers to the views of the group of which one is a member, c)

emotionally.	other party. ”		creates an imaginary opponent • Distorts the standpoint of the other: a) takes remarks out of context, b) oversimplifies by ignoring qualifications or nuances, c) exaggerates by generalization or absolutization The result of the error is to create an imaginary opponent or by taking comments out of context, by exaggeration, by oversimplification.
	• “Rule (4) A party may defend a standpoint only by advancing argumentation relating to that standpoint. ”		*Protagonist* • Advances argumentation that is irrelevant and thus fails to address the standpoint under discussion • Defends the standpoint by using non-argumentative means of persuasion: 1) plays on emotions of the audience, or b) parades own qualities. The result is to defend the argument by using non-argumentative means, e.g., appeal to the audience’s negative or positive emotions or to use one’s expertise, integrity, credibility, etc. as a basis for having their standpoint accepted.

Unexpressed premises must be defended by the protagonist and identified correctly by the antagonist.	• "Rule (5) A party may not disown a premise that has been left implicitly by that party or falsely present something as a premise that has been left unexpressed by the other party. "		*Protagonist or antagonist* • Protagonist denies an unexpressed premise • Antagonist magnifies an unexpressed premise The result is that the premise is not properly addressed.
The starting points of the discussion must be used properly in criticizing and defending standpoints.	• "Rule (6) A party may not falsely present a premise as an accepted starting point nor deny a premise representing an accepted starting point. "		<u>Protagonist or antagonist</u> • Protagonist falsely advances something as a common starting point: a) falsely advances a premise as being self-evident, b) enfolds a proposal in a presupposition of a question, c) conceals a premise in another unexpressed premise, d) advances a circular argumentation (same thing as standpoint) • Antagonist fails to accept a premise represented as a common starting point by casting doubt on it. The result is that the protagonist tries to evade the burden of proof or that the antagonist denies to the protagonist the opportunity to defend the

			standpoint.
The protagonist and antagonist must agree on how to test the soundness of arguments that were not part of the common ground upon starting.	• "Rule (7) A party may not regard a standpoint as conclusively defended if the defense does not take place by means of an appropriate argumentation scheme that is correctly applied. "		*Protagonist* • Chooses an inappropriate argumentation scheme: a) chooses inappropriate symptomatic argumentation, b) chooses inappropriate comparison argumentation, c) chooses inappropriate instrumental argumentation • Uses incorrectly an argumentation scheme that is appropriate: a) uses symptomatic argumentation incorrectly, b) uses comparison argumentation incorrectly, c) uses instrumental argumentation incorrectly. Symptomatic means there is a relation between the standpoint and the premise. The relation may not be valid if based on authority or because "everyone says so". Or the standpoint may be based on non-representative or insufficient observations. Comparison means the relation is one of similarity. An analogy is not correct if its conditions are wrong. Instrumental means the relation is causal. It is used incorrectly if the standpoint is to be rejected because of

			undesirable consequences, a false causal relationship is inferred, or if it is proposed without justification that expected results will worsen a bad situation.
The reasoning underlying the argumentation advanced by the protagonist must be valid. When it is, the standpoint being defended will follow logically from the explicit or implicit premises used by the protagonist. Any unexpressed premises must be made explicit.	• "Rule (8) A party may only use arguments in its argumentation that are logically valid or capable of being validated by making explicitly one or more unexpressed premises. "		*Protagonist* • Confuses necessary conditions with sufficient conditions: a) treats necessary condition as sufficient, b) treats sufficient condition as necessary • Confuses the properties of wholes and parts: a) ascribes structure-dependent or relative property of the whole to a part of the whole, b) ascribes structure-dependent or relative property of a part to the whole. The result is incorrect use of if..then.. arguments or confusing the whole and parts in argumentation.
Misunderstandings can result from incorrect interpretations of vague, unclear, or ambiguous formulations. Misunderstandings can lead to a pseudo-resolution. Premises must, therefore, be expressed	• "Rule (10) A party must not use formulations that are insufficiently clear or confusingly ambiguous and a party must interpret the other party's		*Protagonist or antagonist* • Takes unjust advantage of unclearness: a) uses structural unclearness, b) uses implicitness, c) uses indefiniteness, d) uses unfamiliarity, e) uses vagueness • Takes unjust advantage of

explicitly.	formulations as carefully and accurately as possible. ”		ambiguity: a) uses referential ambiguity, b) uses syntactic ambiguity, c) uses semantic ambiguity.
1.1.1.1.1.5 Stage 4 – Closing			
The protagonist and antagonist determine if the protagonist has successfully defended his or her standpoint. The protagonist must withdraw his or her standpoint if the antagonist’s doubts have not been resolved. If the antagonist’s doubts have been resolved, then the protagonist’s standpoint holds.		• Establishes the result • Accepts or fails to accept, • Upholds the failure to accept the standpoint • Defines, makes precise, amplifies, etc. • Asks for elaboration.	
The protagonist and the antagonist must agree to the result of the discussion. Both parties must agree that the protagonist successfully defended his or her standpoint.	• “Rule (9) A failed defense of a standpoint must result in the party that put forward the standpoint retracting it and a conclusive defense of the standpoint must result in the other party retracting its doubt about the standpoint. ”		*Protagonist or antagonist* • Protagonist makes an absolute of the success of the defense by concluding that his or her standpoint is true merely because it was successfully defended. • Antagonist makes an absolute of the success of the defense by concluding that his or her standpoint is true merely because the protagonist was not able to defend successfully the opposite. The result is double errors

			on both parts. The protagonist incorrectly attributes fact without justification. The antagonist confuses the role with that of the protagonist and assumes incorrectly that if the positive is not proved, then the negative is correct.
Misunderstandings can result from incorrect interpretations of vague, unclear, or ambiguous formulations. Premises must, therefore, be expressed explicitly.	• "Rule (10) A party must not use formulations that are insufficiently clear or confusingly ambiguous and a party must interpret the other party's formulations as carefully and accurately as possible."		*Protagonist or antagonist* • Takes unjust advantage of unclearness: a) uses structural unclearness, b) uses implicitness, c) uses indefiniteness, d) uses unfamiliarity, e) uses vagueness • Takes unjust advantage of ambiguity: a) uses referential ambiguity, b) uses syntactic ambiguity, c) uses semantic ambiguity. The result is that misunderstandings can lead to a pseudo-resolution.
1.1.1.1.1.6 Post-closing - Next step			
A new critical discussion may commence with the closing of the previous one. Alternatively, if the just-completed discussion was embedded in a larger one, the			

argumentation will return to the larger discussion, and the process will begin anew with another sub-argument.			

According to Walton

Walton has also defined a four-step method for examining an argument. The purpose of the method is to confirm that the discussion is contributing to the goal of resolving a difference of opinion. The main points are

- Identify the argument
- Identify the contexts of the dialogue
- Establish burden of proof
- Evaluate the criticisms.

The information in the following table is from Walton (1998, Chapter 10).

1. Identify the argument	Identify the propositions of the reasoning, i.e., the premises and conclusions.
	Determine if the reasoning was deductively valid or invalid.
	Determine if the reasoning was inductively weak or strong.
	Determine if the sub-arguments were directed toward the main goal.
	Identify implicit, missing, or excessive premises.
	Identify ambiguity or equivocation.
	Determine if the conclusion is supported by the sequence of reasoning.
2. Identify the contexts of the dialogue.	Identify the type of dialogue, e.g., critical discussion.
	Determine if the goals of the dialogue established relevance.
	Determine if the issue of the dialogue is supported by the global conclusion.
	Determine if the dialogue type shifted.
	Determine if evidential priority was established.
	Determine if the participants' positions shifted during the dialogue.
3. Establish burden of proof.	Determine the burden of proof.

	If the reasoning is inductive, determine if premises are missing or if inductive reasoning errors occurred.
	If the reasoning is deductive, determine if premises are missing.
	Determine if presuppositions of questions were loaded or multiple.
4. Evaluate the criticisms.	Evaluate criticisms as they are raised. Determine if errors occurred.
<u>General rules</u>	
• Locution rules	Defines the kinds of speech acts, e.g., questions.
• Dialogue rules	Defines turn taking and guidelines for advancing speech acts.
• Commitment rules	Defines the participant reaction to which each type of speech act leads.
Strategic rules	Defines the win-loss conclusion.
• Rules of relevance	Participants must stick to the goal of dialogue or be challenged.
• Rules of cooperativeness	Participants must answer cooperatively. They must not deny their position.
• Rules of informativeness	The speaker must tailor his or her responses to the respondent's knowledge. The speaker must not provide any more than the minimally required information.

12. CHALLENGES AND POSSIBILITIES FOR TRAINING CRITICAL THINKING

It is appropriate now to summarize some of the implications of this theory for the challenges we laid down in Chapter 1. Here again are some of the potential difficulties of implementing critical thinking training in the Army context:

Is Critical Thinking Consistent With Tactical Battlefield Constraints?

- Will critical thinking on the battlefield take too much time? Would that time be put to better use gaining a jump on the enemy?
- Will critical thinking result in a loss of the confidence necessary for decisive leadership and action? Will it undermine the "will to fight"?

The external layer of critical thinking, i.e., the assessment of reliability, is the source of a stopping rule for the process of challenging and response. It demands that the critical thinker stay focused on real task objectives. Reflective reasoning is one tool among others, including recognitional decision making, and should be used when and only when it will increase the odds of success. There are, however, many examples in which a little time spent thinking saved much more time in execution (e.g., Cohen & Thompson, 2001). Because of the external layer, however, critical thinking never involves an endless exploration of alternative possibilities with no end in sight.

The critical dialogue layer of critical thinking permits a variety of different reasoning styles that differ in how free-ranging the consideration of alternative possibilities may be. In time stressed situations, a more constrained reasoning process, in which basic assumptions are not questioned, leads to more rapid decision making. Explicit recognition of the mode of dialogue that has been adopted among team members may actually speed up communication and reasoning. Confidence is typically increased by a disciplined exploration of relevant and significant alternative possibilities.

Is Critical Thinking Consistent With Other Battlefield Skills?

- Will critical thinking skills trump experience or leadership qualities on the battlefield, which might in fact lead to better decisions?
- Will critical thinking be too "critical"? Will it stifle innovation or the development of new tactics and techniques?

The external layer of critical thinking involves choosing the most reliable process for a given decision. For experienced leaders, the most reliable method sometimes involves trust in their own gut feel for a situation.

As far as innovation goes, the dialogue layer of critical thinking is not "critical" in a narrow, negative sense. It not only evaluates possibilities, it stimulates the generation of *new* possibilities. The space of alternatives is constantly changing as a result of the challenge and response process. The construction of these mental models does not necessarily proceed in a rigid step by step fashion. In the context of a permissive critical dialogue, any assumptions may be questioned and retracted. Alternative mental models are evaluated in terms of their overall coherence with a system of beliefs. The interconnectedness of beliefs in a coherence-based system can lead to rapid, creative shifts in the understanding of a situation, similar to the

paradigm shifts that T. Kuhn (1996) describes. Such shifts may involve the simultaneous modification of numerous assumptions, beliefs, and plans.

Is Critical Thinking Appropriate for Military Organizational Structure?

- Will critical thinking encourage inappropriate initiative? Will it disrupt the chain of command and degrade coordination and synchronization on the battlefield? Put another way, is the Army too centralized and hierarchical for critical thinking to flourish?
- Will critical thinking hinder the development of trust in diverse, multi-cultural teams because it is "Western, masculine, individualistic, adversarial, and coldly rational" (Atkinson, 1997; cited in Davidson, 1998).

Critical thinking is most suited to situations in which individuals have significant autonomy and responsibility, and such situations are likely to increase in frequency in future Army missions. But critical thinking can function at many different levels, e.g., in the performance of virtually any non-routine task. The dialogue layer provides a series of dialogue types that vary in the extent to which assumptions are questioned. The higher the level of initiative, the more far-reaching the exploration of alternatives might be. But critical thinking at some level is nearly always appropriate.

As for cultural diversity, the dialogue layer provides a framework for classifying different styles of interaction. This framework may lead to more stable and better calibrated expectations among individuals from diverse cultural backgrounds. It also allows for the evolution of new styles of dialogue that may be better suited to a specific team or context.

Will Critical Thinking Fit into Army Training?

- Are there “right answers” in critical thinking? If so, isn’t this just a new phrase for teaching doctrine and tactics, which we already do? If not, what good are skills that can’t be evaluated? How can we know they will improve performance?
- Will critical thinking instruction consume too much training time? How will we persuade instructors to provide that time? Does critical thinking require technical training in logic or decision theory? Does it require stand-alone courses? How will we persuade students to devote their time to the study of critical thinking?

Metrics for critical thinking performance focus on process rather than product. Both the dialogue layer and the reliability layer evaluate belief acceptance in terms of the processes that led to it, and each provides relatively unambiguous evaluative criteria. Metrics for a successful dialogue measure the degree to which an actual conversational exchange corresponds to the profile of the relevant type of dialogue. For example, was disagreement acknowledged? Were challenges sought out? Were they answered? Metrics for reliability include the probability that the selected cognitive faculty or communicative process will support the objectives of the task under the prevailing conditions. For either dialogue or reliability based measures, a decision may be good even the outcome happens to be bad, and conversely, a decision may be bad even though there was a lucky outcome.

Each layer of critical thinking is associated with a specific set of skills and training objectives. For example, the innermost, mental model layer involves the ability to generate

possibilities based on existing elements, the ability to add dimensions to the space of situations, and the ability to evaluate and compare mental models in terms of their internal coherence and compatibility with background knowledge. The dialogue layer involves awareness of different types of dialogues with different rules for identifying conflicting positions, for challenging and retracting assumptions, and for "winning" and "loosing." The outermost, reliability layer requires an awareness of strengths and weaknesses of different cognitive processes or faculties, and the ability to make appropriate choices based on the circumstances, e.g., between recognitional decision making, creative brainstorming, or reflective reasoning.

Critical thinking skills are best acquired in the context of actual decision making. Thus, critical thinking training may be incorporated relatively seamlessly into subject matter coursework, exercises, and field training. Students may be taught through coaching, hints, feedback, and example, in addition to explicit instruction (see Cohen, Thompson, Adelman, Bresnick, Shastri, & Riedel, 2000b). Critical thinking training can also be given as a standalone course, as long as concrete exercises (e.g., tactical decision games) are emphasized. None of the relevant skills requires specialized training in formal logic, decision theory, or philosophy. Nevertheless, these are skills that need some explicit attention, and thus it would be best for instructors to receive some specialized training. A useful first step might be the development of a brief, intensified critical thinking course for instructors.

APPENDICES

A. GOALS FOR A CRITICAL THINKING CURRICULUM – AND FOR GUIDING ITS ASSESSMENT

Robert E. Ennis

University of Illinois, UC 3/20/2000[117]

Critical thinking, as the term is generally used these days, roughly means reasonable and reflective thinking focused on deciding what to believe or do. In doing such thinking, one is helped by the employment of a set of critical thinking dispositions and abilities, which shall be outlined and which can serve as a set of goals for a critical thinking curriculum and its assessment. Elsewhere, I have elaborated these goals in varying degrees.[118]

Dispositions

Ideal critical thinkers are disposed to:

1. Care that their beliefs be true, and that their decisions be justified; that is, care to "get it right" to the extent possible. This includes the interrelated dispositions to:

    A. Seek alternatives (hypotheses, explanations, conclusions, plans, sources), and to be open to them;

    B. Endorse a position to the extent that, but only to the extent that, it is justified by the information that is available;

    C. Be well informed; and

    D. Consider seriously other points of view than their own.

2. Care to present a position honestly and clearly, theirs as well as others'. This includes the dispositions to:

    A. Be clear about the intended meaning of what is said, written, or otherwise communicated, seeking as much precision as the situation requires;

    B. Determine, and maintain focus on, the conclusion or question;

    C. Seek and offer reasons;

    D. Take into account the total situation; and

    E. Be reflectively aware of their own basic beliefs.

---

[117] Originally presented in July, 1994 at the Sixth International Conference on Thinking at MIT, Cambridge, MA. The current version incorporates minor revisions. A subsequent version has been published and copyrighted by ASCD in Arthur Costa (Ed.), *Developing Minds* (Washington, DC: ASCD, 2001) under the title, "Goals for a Critical Thinking Curriculum and Its Assessment" , on pages 44-46. I reserve a right to grant permission to other persons to copy this work in whole or part in various forms, provided that appropriate notice of permission and copyright accompanies such copying or publication.

[118] Including "A Concept of Critical Thinking," Harvard Educational Review, 32 (1962), pp. 81-111; "A Logical Basis for Measuring Critical Thinking Skills," Educational Leadership, 43, (1985), 2, pp. 44-48; "A Taxonomy of Critical Thinking Skills and Dispositions,: in Joan Baron and Robert Sternberg (Eds.), Teaching Thinking Skills: Theory and Practice (New York: W.H. Freeman, 1987); "Critical Thinking: A Streamlined Conception," Teaching Philosophy, 14 (1991), 1, pp. 5-25; and Critical Thinking (Upper Saddle River, NJ: Prentice Hall, 1996).

3. Care about the dignity and worth of every person. This includes the dispositions to:

A. Discover and listen to others' view and reasons;

B. Avoid intimidating or confusing others with their critical thinking prowess, taking into account others' feelings and level of understanding; and

C. Be concerned about others' welfare.

Comments About These Dispositions

1. Several of the dispositions (1D, 2E and 3A) contribute to being well-informed (1C), but are separate dispositions in their own right.

2. With respect to epistemological constructivism (the view that truth is constructed): in expressing a concern about true belief, this depiction accepts the view that our concepts and vocabulary are constructed by us, but also that (to oversimplify somewhat) the relationships among the referents of our concepts and terms not constructed by us. We can have true or false beliefs about these.

With respect to pedagogical constructivism (the view that students learn best when they construct their own answers to problems and questions): for some (but not all) goals and types of learning, this view has empirical support, but it should not be confused with epistemological constructivism. In particular, the validity of pedagogical constructivism (to the extent that it is valid) does not imply the validity of epistemological constructivism. They are totally different ideas.

3. The disposition to care about the dignity and worth of ever person (#3) is not required of critical thinking by definition, but in order that it be humane. I call it a "correlative disposition," by which I mean one that, although it is not part of the essential definition of 'critical thinking', it is desirable for all critical thinkers to have. The lack of it makes the critical thinking less valuable, or of no value at all, or even dangerous on occasion.

A criticism of critical thinking for a definitional omission of caring for the worth and dignity of every person could well be based on the unreasonable assumption that the concept, critical thinking, should represent everything that is good, an overwhelming requirement indeed. On the other hand, any educational program that includes critical thinking, but not the correlative disposition to care about every person's worth and dignity would be deficient. The power of critical thinking unaccompanied by this correlative disposition could lead to serious trouble.

4. Pedagogical usefulness, not elegance or mutual exclusiveness, is the goal of these lists.

Abilities

Ideal critical thinkers have the ability to:

(The first five items involve clarification.)

1. Identify the focus: the issue, question, or conclusion;

2. Analyze arguments;

3. Ask and answer questions of clarification and/or challenge;

4. Define terms, judge definitions, and deal with equivocation;

5. Identify unstated assumptions;

(The next two involve the basis for the decision.)

6. Judge the credibility of a source;

7. Observe, and judge observation reports;

(The next three involve inference.)

8. Deduce, and judge deductions;

9. Induce, and judge inductions

A. to generalizations, and

B. to explanatory conclusions (including hypotheses);

10. Make and judge value judgments;

(The next two involve supposition and integration.)

11. Consider and reason from premises, reasons, assumptions, positions, and other propositions with which they disagree or about which they are in doubt - without letting the disagreement or doubt interfere with their thinking ("suppositional thinking");

12. Integrate the other abilities and dispositions in making and defending a decision;

(The next three are auxiliary critical thinking abilities: having them, though very helpful in various ways, is not part of the constitutive core of being a critical thinker.)

13. Proceed in an orderly manner appropriate to the situation, for example,

A. follow problem solving steps,

B. monitor their own thinking (that is, engage in metacognition), and

C. employ a reasonable critical thinking checklist;

14. Be sensitive to the feelings, level of knowledge, and degree of sophistication of others; and

15. Employ appropriate rhetorical strategies in discussion and presentation (orally and in writing), including employing and reacting to "fallacy" labels in an appropriate manner.

Further Comments

1. This is only a critical thinking content outline. It does not specify level, curriculum sequence, emphasis, teaching approach, or type of subject-matter content involved (standard subject-matter content, general knowledge, symbolic content, streetwise-knowledge content, special knowledge content, etc.).

2. If this outline is used to organize a separate critical thinking course or module, the definitional and assumption-identification abilities would problem come later than in the given order, because of their difficulty. In any course, whether it be a separate critical thinking course or module, or one in which the critical thinking is infused in or immersed in standard subject-matter content, or some mixture of these, all of the dispositions, the suppositional and integrational abilities (#11 and #12), and auxiliary abilities #13 through #15 should permeate the course.

3. The fallacy-labels part of #15 is partly rhetorical, and partly constitutive of critical thinking. The constitutive parts are covered in #1 through #12, leaving the rhetorical part of #15. These labels, including such terms as “circularity,” “bandwagon,” “post hoc,” “equivocation,” “non sequitur,” and “straw person,” are useful to know, but dangerous when used by, or in the company of, people who do not understand them fully, because the terms are so easy to apply and misapply and, on occasion, are intimidating.

Summary and General Comments

In presenting this brief and fairly abstract list of critical dispositions and abilities, I have only attempted to depict, rather than defend, them. The defense would require much more space than is available, but would follow two general paths: 1) examining the traditions of good thinking in existing successful disciplines of inquiry, and 2) seeing how we go wrong when we attempt to decide what to believe or do.

Although these dispositions and abilities could be provided with examples and more detail,[119] I shall instead now provide an overview even more brief:

The ideal critical thinker is disposed to try to “get it right”, to present a position honestly and clearly, and to care about the worth and dignity of every person; furthermore the ideal critical thinker has the ability to clarify, to seek and judge well the basis for a view, to infer wisely from the basis, to imaginatively suppose and integrate, and to do these things with dispatch, sensitivity, and rhetorical skill. There is much more to say about all this in both curriculum planning and assessment, but, as promised, I shall be brief – and stop here.

[119] See the items in footnote 118.

B. ASSUMPTIONS AND PROBLEMS IN MAINSTREAM INFORMAL LOGIC

What Is Informal Logic?

If reasoning cannot always bring certainty, then what does reasoning accomplish? How can we tell good reasoning from bad? Informal logic, like contemporary foundationalism, tries to fill the gap left by the fall of classical foundationalism. Informal logicians do not reject the idea that argument is central to reasoning, only the idea that arguments must be airtight to be acceptable. Thus, informal logicians have tried to define what it means for evidence to be *acceptable* (though not indubitable) and for conclusions to *fit* that evidence (even though not deductively implied by it). While classical foundationalism is relentlessly normative and analytical, informal logic focuses at least in part on *description* of how reasoning is actually conducted. According to Fisher (2000: p. 109), informal logic "…studies 'real arguments' – arguments which are or have been used with the aim of convincing others of a point of view – and it tries to understand and explain what makes such arguments succeed or fail in convincing their audiences in real contexts and what *should* do this… (italics in original). Only in the very last phrase of this quote, does Fisher allude to the ultimate normative purpose of informal logic.

Others, such as Johnson (2000: p. 119), put the normative component in a more prominent position: Informal logic is "…the branch of logic whose task is to develop non-formal$_2$ standards, criteria, procedures for the analysis, interpretation, evaluation, critique and construction of argumentation in everyday discourse… ." In subsequent discussion, Johnson expands the coverage of informal logic to all natural language argument, not just "everyday" argument. Thus, it would include argument in specialized technical fields such as science, law, and medicine.

An issue upon which informal logicians do not agree is the role of formal logic. This disagreement seems symptomatic of a deeper uncertainty about the guiding principles of the field itself. Johnson and Blair (1994, p. 11; Johnson, 2000, pp. 119-120) say that informal logic can be formal in the sense of developing systematic and rigorous theories. On the other hand, it is non-formal in the sense that it rejects syntactically based or proof theoretic criteria of normative adequacy. In the latter sense, it excludes formal logic by definition. What is the basis for this rejection? It seems odd to prejudge the results of research on real-world argumentation, since it might still turn out that formal argument patterns are used in some instances of real world reasoning (e.g., Rips, 1994; although others disagree, e.g., Johnson-Laird & Byrne, 1991). What underlies this prejudgment, perhaps, is the tendency to confuse rules of logical implication (such as *modus ponens*) with strategies of reasoning. Formal rules of implication might serve as criteria of valid arguments, even though they are not procedural recipes for creating such arguments. Thought processes do not necessarily follow the steps of a proof. For example, a deductive argument might be constructed by backwards reasoning from the conclusion, from the premises forward, or from the middle forward to the conclusion and back to the premises, and so on). Even though deductive logic fails to describe the "laws of thought," it may nonetheless supply some of the ingredients for a complete cognitive theory.[120] Johnson (2000: p. 142) concedes that

[120] As Barth herself (1987: p. 35) puts it in discussing the distinction between different senses of *formal*: "The primary obstacle is not 'formality' in whichever sense of that word but …the idea that logic itself is essentially tied to the deductivist conceptions of science…" Deductive logic (whether syntactically or semantically based) is not an

formal logic may sometimes be useful in describing real-world arguments, and thus, that the same researcher might have to be both an informal and a formal logician. The more inclusive view of informal logic, as the general study of real-world arguments, is overridden by the limiting connotations of the word "informal." [121]

Informal logic has developed some distinctive tools and concepts of its own. In textbooks on informal logic, there are three widely accepted and closely interconnected tools for evaluating arguments:

1. Diagramming argument structure, to identify the components of the argument (e.g., reasons and conclusions) and the relationships among the components.
2. Applying criteria to decide the cogency or soundness of the argument, in particular, determining the *acceptability* and *relevance* of the premises. and the *sufficiency* of the inferential link between the premises and the conclusion.
3. Identifying fallacies that might undermine the cogency or soundness of the argument. Fallacies have been classified according to whether they involve violations of acceptability, relevance, or sufficiency (Govier, 1997; Johnson & Blair, 1994).

The first technique, diagramming, is part of argument analysis. The second and third techniques are part of argument evaluation, that is, once the reasons and the conclusions in an argument have been revealed by analysis, the second and third techniques determine whether the reasons in fact justify the conclusion.

According to informal logic, arguments are the key to reasoning. Deciding whether a belief fits the evidence involves evaluating the argument for that belief. Arguments in informal logic are viewed more concretely than they are in formal logic. They incorporates reference to an actual *claim* made by a *person* with a *goal*, i.e., to persuade another person to accept the claim (see the second row of Table 2). The criteria of evaluation include the acceptability of the premises *for the person being persuaded*. Moreover, informal logicians focus on non-demonstrative inferences, in which the premises do not establish the conclusion with absolute certainty. Thus, there is room for variation in the standards of acceptability depending on the domain or on what is at stake in a particular context.

Despite its empirical motivation as a description of real argumentation, informal logic inherits many of its assumptions from contemporary foundationalism. In particular, the methods proposed by informal logic echo the foundationalist notion that a belief is justified either when it needs no argument (i.e., basic beliefs) or when arguments ground it in beliefs that need no argument. Informal logic also echoes a less obvious foundationalist assumption: that arguments for individual hypotheses warrant the addition of new beliefs one at a time. This assumption is important because it leads informal logicians to neglect the need in many critical thinking contexts to construct and evaluate an overall account (or mental model) of the situation rather than infer a conclusion at the end of a series of arguments. Despite its emphasis on uncertainty,

---

adequate overall framework for inference even in technical discourse (Govier, 1987). On the other hand, deductive logic might serve useful but limited functions in everyday discourse, e.g., as a type of sub-dialogue (Walton & Krabbe, 1995).

[121] Others, both outside and inside the field of informal logic, would prefer to restrict the word *logic* to deductive logic (e.g., Hintikka, 1999; McPeck, 1994). The intended sense of logic in the name *informal logic* includes *strategies* for reasoning, not (only) *rules of implication*.

some of the problems with informal logic reflect a continuing submerged influence of formal deductive logic. Unfortunately, simply adding a probabilistic veneer to handle uncertainty does not eliminate that influence.

The failure of mainstream informal logicians to adopt an integrative perspective may have prevented them from providing more effective support for the field of critical thinking. As it is now, critical thinking textbooks borrow eclectically from traditional formal treatments of deductive logic and probability, and from informal logic ideas on structuring and analyzing arguments. These are typically presented in separate chapters. No synthesis is available that is firmly anchored in research on real-life argumentation.

What Is the Structure of an Argument?

Beardsley (1950) pioneered the method that has become standard for analyzing and diagramming the structure of an argument. He puts the rationale for the method this way: "The essence of an augment is that it makes a claim upon belief and supports this claim with a reason or reasons. To find out whether the reasons are good ones, you must take the argument apart and examine it piece by piece." In many cases, however, an argument "is not so simple or so orderly that all relations can be perceived at once." A method is needed for discerning its logical structure. Beardsley specifies three steps in such a method, as indicated in this quote (p. 18):

1. Read the argument carefully.[122]
2. Break it down, by bracketing and numbering all separate statements. Then find and circle words that indicate logical relationships. Some of these indicate a conclusion (such as *therefore, so,* and *I conclude that*), and others indicate a reason (such as *because, for, since,* and *as shown by*). Finally, supply indicator words that are suggested but omitted.
3. Create a diagram which sets out the numbered statements and uses arrows to show which statements are reasons for which other statements.

Simply by reflecting on the possible topologies of such diagrams, Beardsley distinguished three types of argument structure (as shown in Figure 34):

> In a *convergent* argument, several independent reasons support the same conclusion…. In a *divergent* argument, the same reasons supports several conclusions…. A *serial* argument contains a statement that is *both* a conclusion *and* a reason for a further conclusion (p. 19; italics in original).

A particular argument can involve any combination of the three structures. Beardsley suggests some rules of thumb for constructing clear verbal or written arguments based on these ideas. For example: Mention all the reasons for a particular conclusion as close together as possible, and in a serial argument, move in a single direction of inference. This method of organizing discourse by evidence and conclusion contrasts sharply with a *narrative* method of organizing information, in which events are described in a temporal and/or causal sequence in order to provide a complete account, or story, of a set of events (Hastie, 1993; Schum, 1994).

---

[122] The emphasis on written rather than vocal argumentation is characteristic of formal logic, informal logic, and critical thinking. Dialogue logic has broken from this tradition, in emphasizing speech acts in argumentation.

Beardsley has introduced several assumptions inherited by informal logic from formal logic: (i) rational persuasion is best accomplished by presenting information organized in the form of arguments, (ii) analysis is necessary to evaluate an argument, and (iii) argument analysis requires breaking the argument down into small propositional components.

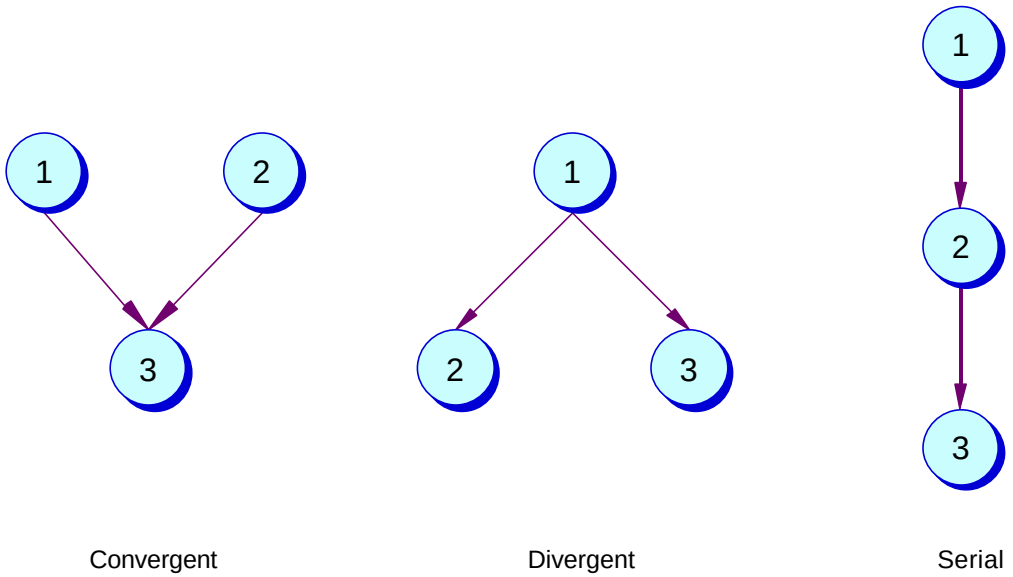


Figure 34. Three types of argument structure identified by Beardsley. Numbers refer to statements. Arrows represent relationships of evidential support.

Thomas's textbook (1997/1973), which deals with what he calls "natural logic" and critical thinking (p. xi), includes some advances on Beardsley's method, and has had a major influence on subsequent textbooks. Beardsley assumes that the *statements* within an argument correspond to the *reasons* for believing the conclusion. In particular, Beardsley recognized only one way in which multiple premises can be relevant to a single conclusion, corresponding to the *convergent* diagram in Figure 34, where each premise corresponds to a reason. We might call this a topological approach, in which the only relationship between two statements is one of being connected or not connected, without discriminations among types of connection. Thomas goes beyond topology, and distinguishes several different ways multiple premises can be relevant to a conclusion. First, arguments with multiple premises may be linked rather than convergent (Figure 35):

> When a step of reasoning involves the logical combination of two or more reasons, they are diagrammed as *linked*… only one arrow is used, to show that the conclusion is a single inference from the combination of both reasons. Reasoning is linked when it involves several reasons, each of which needs the others to support the conclusion…
>
> When two or more reasons do not support a conclusion in a united or combined way, but rather each reason supports the conclusion completely separately and independently of the other, the reasoning is *convergent*… A convergent argument is equivalent to separate arguments…for the same conclusion… the support given

to the conclusion by each separate reason, or line of reasoning, would remain the same even if the other … reason(s) were false (pp. 50-53).

The introduction of linked arguments provides the basis for filling in missing premises. Since "the author may have failed to set down all the assumptions," one may notice a "hole where an essential part of the structure is missing." Such holes must be filled by means of "personal, logical insight" which supplements the otherwise "mechanical procedures" for diagramming (Thomas, pp. 469-470). Such holes can be noticed only in a linked argument, where the stated premises may "need" some further, unstated premise in order to support the conclusion. Thomas recommended that missing premises be included in diagrams, with brackets to indicate that they were not stated by the author of the argument (Figure 35).

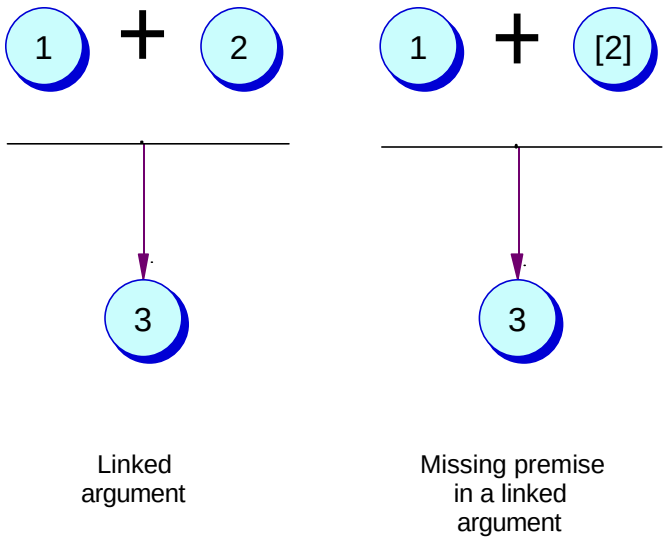


Figure 35. Thomas distinguished linked from convergent arguments, and suggested that missing premises in linked arguments be supplied.

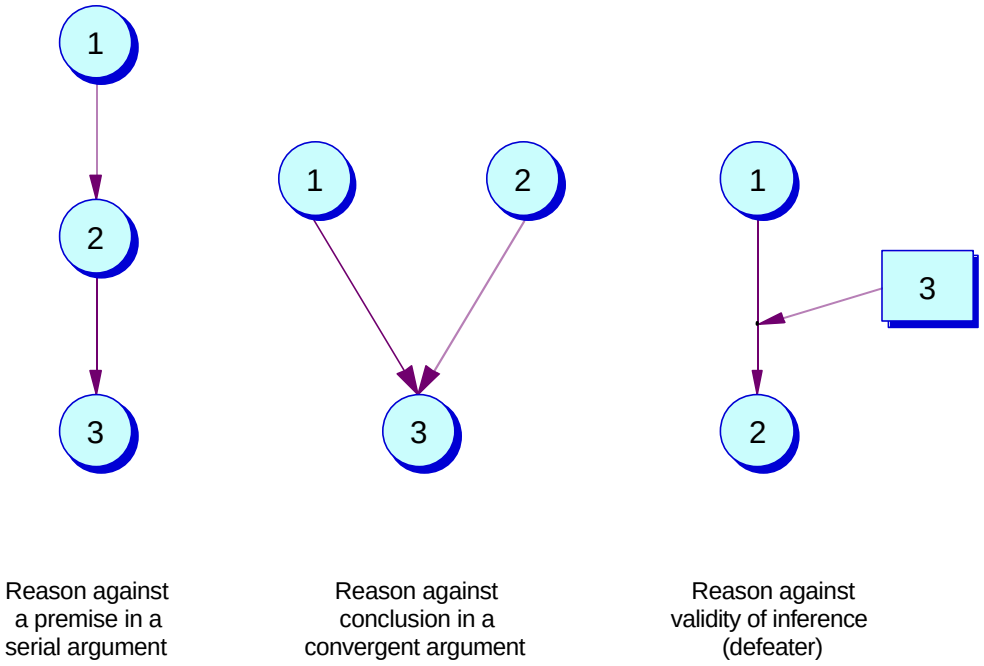


Figure 36. Three kinds of negative reasons distinguished by Thomas.

Thomas went beyond Beardsley's topological approach in a second respect by devoting attention to different ways that reasons might tell for or *against* an argument (Figure 36). First, they might serve directly as reasons against the conclusion. Thomas suggested showing reasons both for and against a conclusion in the same convergent argument diagram, using dotted arrows for the reasons against (chapter 5-1). Second, there might be reasons against one of the premises in an argument; here, dashed arrows can be used in a serial argument, pointing toward the premise that is argued against (chapter 5-1). The third and most interesting category that Thomas identified were reasons against neither the premise nor the conclusion, but against the *validity of the inference step*[123] from a reason to the conclusion (Thomas, chapter 5-2). Reasons in this role are neutralizing *defeaters*, i.e., additional information that overrides the effect of an argument without providing support for the opposite conclusion. Defeaters of an argument describe possible situations in which the premises are true and the conclusion is false. Thomas suggested that defeaters be diagrammed in a distinctive way, as had Toulmin (1958), by drawing dashed arrows from the defeater to the arrow representing the defeated inference step (Figure 36).

Another approach, however, is to treat defeaters (actually, the negations of defeaters) as premises in a linked argument. They certainly fit Thomas' definition of linked premises, since the negations of the defeaters "work together" with the other premises to support the conclusion. If a defeater is true, the support given to the conclusion by the other premises decreases, and may disappear altogether. Earlier, Scriven (1976) had taken precisely this view, and it still perhaps the more usual approach not to separate defeaters out for special treatment. Nevertheless, there is a compelling reason for representing defeaters as Thomas proposed. There may also be defeaters

[123] Many informal logicians avoid use of the term *validity* because of its association with deductive logic, and prefer to use terms like *cogency* (Govier, 1987) or *soundness*. It seems clear enough, however, that Thomas (and many others who chose to use *validity*) do not intend it that narrowly.

of defeaters and defeaters of defeaters of defeaters, and so on. In this case, Scriven's approach would have us to include the negations of odd numbered defeaters and the affirmatives of even numbered defeaters as linked premises in the argument. But this completely obscures the internal structure of the defeaters. In other words, the important relationship of *p defeating q* is not explicitly represented.

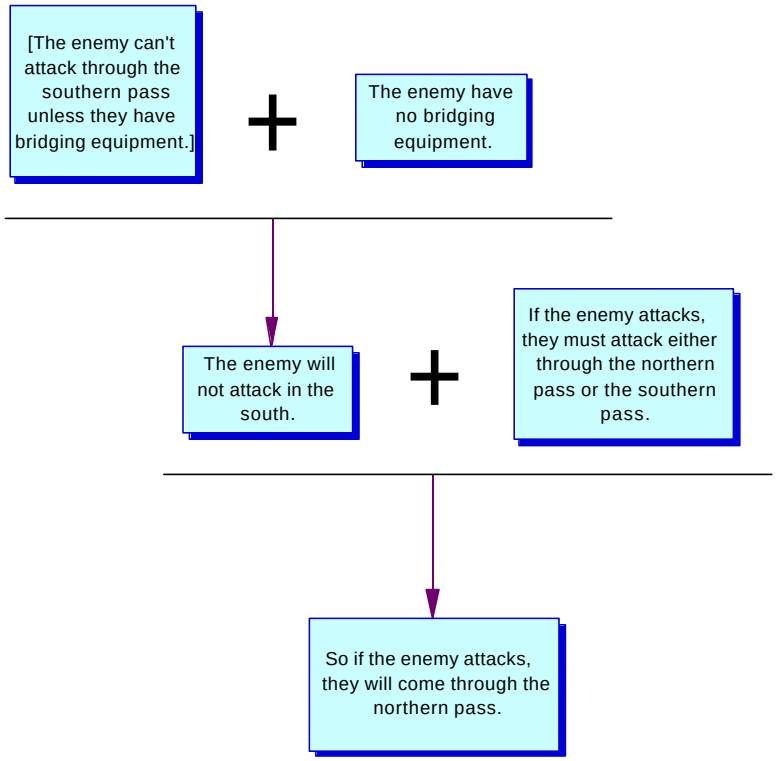


Figure 37. Diagram of B's deductive argument that the enemy will attack through the north (if it attacks), as a serial argument with two linked components and one missing premise.

Let's see how these tools might be used to clarify an example. Figure 37 shows a diagram of B's reasoning as we reconstructed it earlier, in the form of two deductive arguments, which may be read from top to bottom. The first deductive argument requires that we supply a missing conditional premise (shown in brackets). Each of the two arguments in the series is linked (as indicated by the plus signs), which reflects the way premises in a deductive argument "work together" to support the conclusion. The two arguments together form a serial argument, because the conclusion of one is a premise for the other.

Figure 38 diagrams A's counterargument, based on the absence of artillery in the north. We could have supplied a missing conditional premise to make this a deductive argument (e.g., *if the enemy does not artillery in a sector, they will not attack in that sector*). We chose not to because of the willingness of the parties to consider defeating information. If there were a conditional implicit in this argument, it would be under constant revision. Moreover, once revised, it could not be used as the default assumption for reasoning with incomplete information in another situation (see previous discussion).

B now objects that the enemy may have developed longer range artillery. We have followed Thomas' recommendation and diagrammed defeaters, such as B's objection, as distinctive functional elements in the argument. The alternative, as recommended by Scriven and

others, would be to represent A's argument as having two linked premises, one of which is the negation of B's objection:

P1. *Enemy has no artillery in the north.*

P2. *Enemy has not developed longer range artillery.*

C. *Enemy will not attack in the north.*

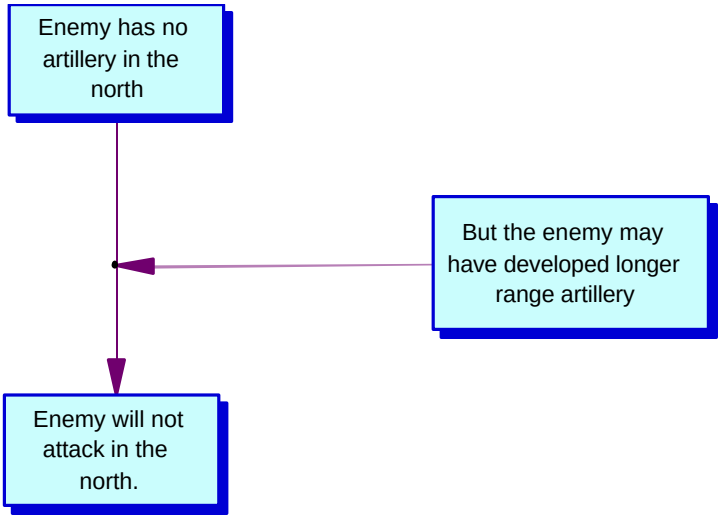


Figure 38. Diagram of A's argument that the enemy will not attack in the north, with B's objection shown as a defeater.

We have already seen one reason not to represent defeaters as linked premises, based on the back and forth nature of challenge, challenge to challenge, and so forth. Deeper insight, and a second reason for treating defeaters differently, comes from viewing argumentation within the context of a dialogue (e.g., Freeman, 1991, chapter 6). The participants in the dialogue incur commitments through their statements or other actions, and these commitments imply a burden of proof (Walton & Krabbe, 1995). The proponent has accepted a burden of proof with respect to the conclusion; her job is to convince the opponent to accept the conclusion. The premises she puts forward are intended to fulfill that burden, and thus she also has a burden of proof to show that the premises are acceptable. Rebuttals, by contrast, are potential objections for which the proponent does *not* have a burden of proof until they are actually raised as objections by the opponent. Indeed, it would be impossible for the proponent to show that no defeating condition was the case, since the list is potentially endless. Thus, if negations of defeaters were premises, no conclusion could ever be accepted. (For example, the absence of artillery in a sector could not be used to argue against intent to attack in that sector unless the proponent could actually *show* that the enemy had not developed longer range artillery.) Rather, it is up to the *opponent* to provide some reason for suspecting that a defeating condition exists. Only then is the burden of proof shifted to the proponent to show either that the defeating condition does not exist or that the conclusion may still be true.

A and B respectively have now given arguments with opposing conclusions regarding location of attack – one by A against the north (Figure 38) and one by B supporting the north (Figure 37). To make a decision, we must evaluate each argument and then, somehow, compare them. Informal logicians (e.g., Govier, 1987) argue for one more step of analysis before

evaluation begins. They suggest combining the two arguments into a single converging argument, as we have done in Figure 39. The rationale for this follows from the assumption that argument evaluation is ultimately a matter of weighing independent reasons. First, if we had made this example longer (and more realistic), both A and B might have presented more lines of reasoning than the single one each produced. These would provide separate, independent grounds for the conclusion; thus, both A and B would have presented converging arguments for their respective positions. To evaluate the strength of each converging argument, we would somehow have to aggregate the strengths of the different independent reasons it contained. But either A or B might also have anticipated objections, or reasons against his position, and included those as *negative* reasons in his argument along with the positive ones. So, the evaluation of each individual argument would have to include weighing of positive and negative reasons. Of course, each would claim that the positive reasons in his argument outweighed the negative reasons in his argument. But this question is precisely what we will have to determine when we compare A's argument and B's argument with one another. Comparing A's argument with B's argument is exactly the same process as evaluating the weight of a single argument that has positive and negative components. For simplicity and consistency, then, it makes sense to place all reasons that bear on the issue, both positive and negative, in the same diagram. By representing them as converging, we indicate that each must be assessed as an independent source of evidence and, ultimately, weighed against the others.

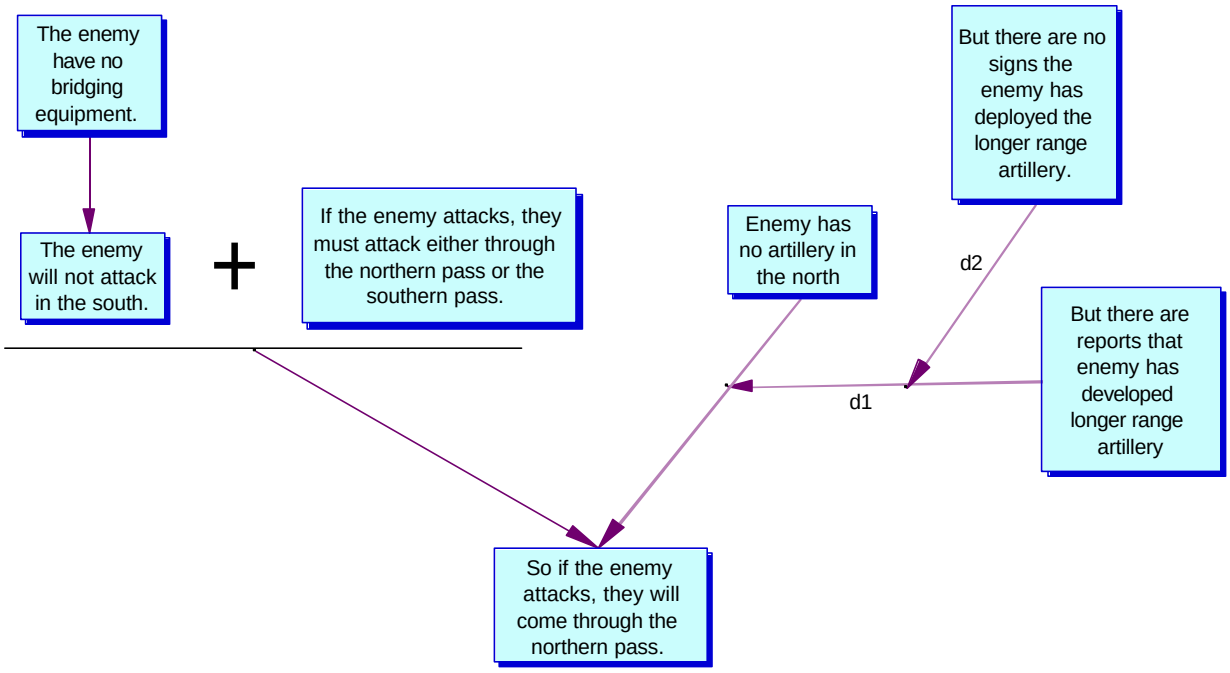


Figure 39. Diagram showing both lines of reasoning in a single argument.

Figure 39 shows A's reasoning and B's reasoning as part of a single argument converging on the conclusion, that the enemy will attack in the north.[124] Since A argues against that

[124] In this diagram, we have dropped the notion that B's argument regarding bridging equipment was deductive, along with the need to supply a missing premise. The missing premise didn't add anything, since it merely restated the inference itself, and presumably this argument might be subject to defeaters, although none have been mentioned.

conclusion, the link from A's premise is shown as a dashed line. B's objection to A's argument is also shown as a dashed line, labeled d1. A's response, that there are no signs that long range artillery has actually been deployed, is also graphed as a dashed line (labeled d2), to show that it is directed against the *relationship* between B's defeater and A's original argument. The numbered labeling illustrates a rule of thumb: In a series of defeaters, defeaters of defeaters, etc., even-numbered reasons support the original argument. Thus, if a series comes to an end at an odd number, there is an unanswered objection to the original claim.

Argument analysis, including diagramming, is meant to be a preparatory step before evaluation of the argument.[125] Nevertheless, the diagramming process itself presupposes the recognition of inferential relationships among statements, which is in part a matter of evaluation. For Beardsley's simple topological approach, the role of such evaluative judgments is minimal. But the distinctions that Thomas introduced –linked vs. convergent vs. defeaters – are more subtle and as a result, more controversial. There is disagreement on whether the distinction between linked and convergent holds up at all (Walton, 1996b) and on the need to supply missing premises in linked arguments. Many if not most writers have ignored the role of defeaters as a separate argument component. But let us now take a look at argument evaluation itself.

What Are the Criteria of a Good Argument?

Informal logic texts depict argument evaluation as process consisting of two independent steps: "There are two essential aspects of good arguments: (i) acceptable premises and (ii) a conclusion that follows from these premises" (Groarke, Tindale, & Fisher, 1996). Similarly, according to Govier (1997: p. 74), "The basic elements of a cogent argument …are as follows: 1. Its premises are all acceptable…2. Its premises are properly connected to its conclusion…" There is also considerable agreement on how these steps are accomplished, once argument analysis, or diagramming, is complete. A popular approach is to break the second step into two parts: assess the relevance of the premises considered individually, and then assess the sufficiency of the entire set of premises as support for the conclusion. Johnson and Blair (1994) introduced these ideas in the form of a three-part classification of reasoning fallacies: fallacies of *relevance*, of *sufficiency*, and of *acceptability* (p. 55). Later in that textbook, they proposed the same three criteria as a general framework for evaluating arguments, and suggest that they be applied in a series of discrete steps:

> First, look at the main premises: P1, P2, P3…C. Ask yourself, Are these premises acceptable? …Remember that this requirement applies mainly to premises that stand alone, without any support… Next, check the premises individually for relevance. Ask, Does P1 satisfy the relevance requirement? Then repeat the process for P2, P3, and so on. Finally, check for sufficiency: Do the premises, taken together…provide enough support for the conclusion? Can you think of a way in which all the premises could prove to be relevant and acceptable and yet

[125] Scriven (1976), another pioneer of diagramming technique, identified seven steps of argument analysis (p. 39). The preliminary steps include: Clarification of meaning of the argument and its components; identification of stated and unstated conclusions; diagramming of structure; formulation of unstated assumptions or 'missing premises'. The evaluation steps include: Criticism of the given and 'missing' premises, criticism of the inference; introduction of other relevant arguments; and overall evaluation of this argument in the light of the above.

the conclusion not acceptable? Is there evidence that you need to have, but do not have? (pp. 268-269).

This scheme, or close variants of it, has been widely adopted in informal logic and critical thinking texts (e.g., Govier, 1997; Freeman, 1993; Groarke, Tindale, & Fisher, 1997). Govier (1987) calls the three criteria the ARG conditions, for acceptability, relevance, and good grounds. Johnson and Blair do not precisely define the three crucial concepts, but here is how Govier (and others) explained them (p. 74; italics added):

1. Acceptability of premises: "…it is *reasonable* for those to whom the argument is addressed to believe these premises."
2. Relevance of a premise to the conclusion: it "give(s) at least *some evidence in favor* of the conclusion's being true."
3. Sufficient or good grounds for the conclusion: "…considered together, the premises give *sufficient reason* to make it *rational* to accept the conclusion…."

These criteria are in part empirical hypotheses about the types of fallacies that are commonly recognized in reasoning (Govier (1997: p. 74; Johnson & Blair, 1994). Different fallacies or types of errors are associated with each of the criteria. For example, a premise might be judged *irrelevant* for a variety of different reasons: e.g., because it attacks the personality or character of a person in order to refute her beliefs (the *ad hominem* fallacy), because of a fallacious *appeal to ignorance* (not disproved, therefore true), because it attacks a false representation of the opponent's position (*straw man* fallacy), because the conclusion "doesn't follow" (*non sequitur*), and so on. Premises may be judged unacceptable or an argument may be judged insufficient for many different reasons as well.

But these definitions do not satisfy the condition that criteria of normative adequacy be unambiguously descriptive. (i) There is no descriptively unambiguous criterion for determining when all the relevant information has been considered. Sufficiency is especially problematic because of defeasibility. In everyday arguments evidence that appeared to be sufficient can always be undermined by unexpected new information. (ii) The definitions use normative expressions like *reasonable to believe*, *evidence in favor*, and *sufficient reason to make it rational to accept*, which are in the same boat as the terms that they define *(acceptable*, *relevant*, and *sufficient support*). They are not likely to provide enough guidance to guarantee consistent judgments by different evaluators. Definitions of specific fallacies may help a little, but also cannot be applied without considerable use of judgment regarding the context in which the alleged fallacy has been committed (Walton, 1989).

(iii) A final component of descriptiveness is how the various criteria are combined to produce a judgment about the cogency of an argument. In the case of formal logic, either inconsistency or invalidity signaled complete failure of the argument, so it was reasonable to regard both as necessary. Can we say the same here, i.e., that an argument is normatively adequate only if every inference step is sufficient, and every premise is relevant and acceptable? In non-deductive reasoning, this is not so plausible. Multiple lines of reasoning may be combined in a single complex argument to establish a conclusion with confidence. Some of these may be "fallacious" while others are not. So it must be determined if the problems that have been identified are bad enough or pervasive enough to spoil the entire argument. This in turn requires a sophisticated effort of *argument structuring*, so that different, independent lines of reasoning

can be distinguished and evaluated separately. Unfortunately, principles of argument structuring, to determine when and how premises are "linked" in the relevant sense have proven difficult to formulate (Walton, 1996b).

Another way to shed light on the three ARG concepts and the roles that each of them plays in a critical evaluation process, is in terms of *probability*. Informal logicians, like contemporary foundationalists, do not expect an argument to provide certainty regarding its conclusion. As we saw earlier, justification, from the internalist point of view, is *evidence that makes a conclusion sufficiently probable*. Thus, the risk of error must be low, or at least, low enough for the context. A critic poses the question *How do you know?,* the proponent answers with information that is intended to increase the expectation that the conclusion is true (i.e., reasons), and then the critic evaluates the answer in term of its impact on the probability of the conclusion.

The ARG concepts can be understood in a qualitative way in terms of probability conditional on evidence and an acceptance threshold:

- A belief P is *relevant* to another belief C just if it makes a difference to the probability of C, i.e., the probability of C given the truth of P is greater than or less than the probability of C given the falsity of P.
- A set of beliefs is *sufficient* for a conclusion C just if the probability of C given the truth of the beliefs is greater than some threshold for acceptance.[126]
- A belief C is *acceptable* for a person just if it is a basic belief or else the person has a set of premises each of which is itself *acceptable* to her and which are individually *relevant* and jointly *sufficien*t for C.

Informal logic thus provides an analysis of a relatively complex concept, a cogent or normatively adequate argument, into three not so simple components: acceptability, sufficiency, and relevance. Probability concepts enable us to reduce these in turn to two more elementary and more general concepts, conditional probability and a threshold of acceptance, and to illuminate the role the three ARG concepts play in argument evaluation.[127]

This does not, however, give us a *descriptive* criterion of argument adequacy. Probability, as it is used in this context, is itself normative. It represents *rationally warranted degrees of belief based on evidence*, or epistemic probability, not statistical frequencies or subjective degrees of belief (Pollock, 1995). Epistemic probability is convenient for some purposes at the

---

[126] An inference might be sufficient according to this definition but fail because of defeaters. If sufficiency is meant to exclude the possibility of defeasibility, we must add a condition such as the following: *…as long as there are no additional beliefs (i.e., defeaters) such that conjoining them to the original premises lowers the probability of C below the threshold.* But this isn't quite adequate, since an inference might be insufficient according to the revised definition but in fact succeed because of defeaters to the defeaters. So, we need to add another clause: *…as long as there are no additional beliefs such that conjoining them to the original premises and the defeaters raises the probability of C above the threshold.* These two clauses allow an argument to be sufficient even if there are potential defeaters (as long as there are defeaters of the defeaters) that have not been made explicit. But the evaluation of the argument is no longer based on internally accessible cues. The account has become externalist.

[127] Here is an (overblown) analogy which captures the logic of this strategy. Biologists learned to characterize an organism in terms of phenotypic features, and then to explain complex phenotypes in terms of simple genotypic entities. The latter was progress even before genes were explained via DNA.

normative level, but it does not supply the missing descriptive criteria. Informal logic takes the relationship of *justification by evidence* as itself a basic, unjustified assumption.

Do the Criteria Apply to Process Instead of Product?

A response to the objection that the ARG criteria are not descriptive is to reinterpret their role. Instead of viewing them as criteria for the evaluation of argument as a *product* (Is this argument sound? Does it establish its conclusion?), we might see them as criteria to be used by evaluators for assessing the performance of critical thinkers (Were the right questions asked about the argument? Were they answered? Were the answers taken into account in follow-up questions or in conclusions?). While not purporting to provide unambiguous criteria for argument quality, ARG might still be seen as an unambiguous proposal for how the *process* of argument evaluation should be conducted. For example, Govier (1997: p. 79), and Johnson and Blair suggest that the criteria be applied in a particular order: *Are the premises acceptable – Are they relevant – Are they sufficient*. On this view, there is a compound of three parts: an argument conceived as a finished product, someone thinking critically about that product who poses ARG questions in a particular order to evaluate it, and an evaluator of the critical thinker who makes sure that the ARG questions are properly posed. ARG provides guidance, but not descriptive criteria, for the critical thinker. But it does provide descriptive criteria for the evaluator.

Freeman (1991) has taken the theme of questioning much further. Instead of focusing on the argument as a finished product, he focuses on argumentation as a process, and incorporates critical thinking as an essential part of the process by which an argument is constructed over time. Following Toulmin (1958), Freeman sees argumentation as a dialogue between a proponent and an opponent, or critic. We imagine that the proponent has made a statement and the opponent has asked a question: *Why should I believe that?* The answer to that question is a *simple* argument: i.e., a reason for believing the conclusion. Freeman (1991: pp. 38-39; 1993: p.85) presents four categories of *basic dialectical questions* that the opponent (or critical thinker) can now ask about that simple argument:

(I) Acceptability question: *How do you know that reason is true?*

Answer: Serial argument structure

(II) Relevance question: *Why is that reason relevant to supporting your claim?*

Answer: Linked argument structure

(III) Sufficiency questions:

1. *Given your reason, how confident should I be of your claim?*

Answer: Modal qualifier (e.g., *probably*, *certainly*, etc.)

2. *Can you give me an additional reason?*

Answer: Convergent argument structure

3. *Why do your premises make you so sure (in light of condition or counterevidence R)?*

Answer: Rebuttals (possibilities that would neutralize argument)

Question (I) asks about the *acceptability* of the reason: What are the reasons for believing it? The answer is a *serial* argument structure (Figure 34) in which the premise of the original argument

becomes the conclusion of a newly supplied argument. In a serial argument, the reasons for accepting the premises are made explicit. Question (II) asks about the *relevance* of the reason. The answer is a *linked* argument, in which new premises are supplied that work together with the original premises to imply the conclusion. In a linked argument (Figure 35) each premise helps explain how the other premises support the conclusion. Question (III) asks the *degree* to which the current reasons are sufficient for establishing the conclusion. This leads to an argument component that represents the force of the argument (e.g., words like *certainly, perhaps, probably*). Freeman (like Toulmin) calls this a *modal qualifier,*. If the reasons are not *sufficient* to establish the conclusion, Question (IV) is appropriate. It requests additional independent reasons for believing the conclusion. The answer is a *convergent* argument (Figure 35), in which reasons contribute independent weight to the force of the argument. Convergent reasons are necessary when no one reason (or linked set of reasons) is sufficient to establish the conclusion on its own. Finally, Question (V) mentions a possible counterexample to the argument, a situation in which the argument would not go through. The proponent might anticipate or acknowledge the objection by including rebuttals (i.e., defeaters) in the argument structure, indicating limitations of scope of the conclusion. If the opponent produces evidence that a rebuttal is true and does apply to the intended scope of the conclusion, the proponent must either defeat that evidence or show how the argument can still work, by providing a counterargument against the rebuttal (a defeater of the defeater).

For Toulmin and Freeman, form follows function. The solution to the problem of argument structure is in the types of questions a rational judge can ask. We determine if the answers "function differently in the economy of arguments, so that different structures appropriately picture their different functions…" (p. 32). Each element of the argument structure is associated with a distinct question-and-answer exchange.

The goal is to provide a vocabulary and a structure for argument evaluation. This contribution, even in the absence of descriptive criteria for good arguments, might be expected to reduce inconsistency in argument evaluation simply by directing attention and facilitating communication. Of course, from an external point of view, the ability of this particular vocabulary and this particular sequence of questions to facilitate communication and consensus should be tested by means of experiment. In any case, it is clear that the three concepts (acceptability, relevance, sufficiency) are more satisfactory at least in principle as criteria applied to *process* than to product. It is a *descriptive* fact that certain questions were or were not asked about an argument, that the questions were or were not answered, and that the answers were or were not taken into account. These facts could serve as the basis for an evaluation of the critical thinking process without our being able to determine whether the questions were answered *correctly*. As evaluators of critical thinking, we do not need to know whether the premises were *actually* acceptable, relevant, and sufficient.

The shift to a dialogue context has promise for the development of more descriptively unambiguous criteria. But we need to look closely at the kind of process guidance that is given. Unfortunately, the standard informal logic criteria are flawed, even when regarded in this light, as guidance for the process of dialogue and reasoning. We will explore some of the problems in the remainder of this chapter.

Will the Criteria Work?

As we have seen, informal logic makes a hard distinction between criteria that apply to premises (acceptability) and criteria that apply to the link between premises and conclusion (relevance and sufficiency). In particular, sufficiency is not meant to refer to the adequacy or cogency of the argument as a whole. Sufficiency (like relevance) is independent of the acceptability or truth of premises; sufficiency of the inference and acceptability of the premises are separate hurdles that an argument must clear in order to be cogent.[128] To assess sufficiency, we temporarily *assume* that the premises are true, regardless of whether or not they are true or even plausible. Premises are sufficient to establish a conclusion if the conclusion is probable *given that the premises are true*. While sufficiency pertains to the set of premises as a whole, relevance pertains to individual premises, but the same principle applies. A premise is relevant if the probability of the conclusion is different when we i*magine the premise to be true* from what it is when we *imagine the premise to be false*. Again, it makes no difference whether the premise really is true or plausible.

An important consequence of separating premise evaluation from inference evaluation is that the argument evaluation process can be divided into separate stages. The cognizer may first challenge the relevance or sufficiency of a set of premises. Then, if satisfied, she turns to the acceptability of the premises. Challenging acceptability may result in the generation of a serial argument structure, in which a further inference is used to derive the original premises. This new inference can now be investigated for relevance and sufficiency, and then, if satisfied, the new premises may be examined for acceptability, and so on. The definition can be applied iteratively to generate a serial argument, but since it is recursive, it needs a closure condition. That is, the series of arguments must be grounded at some point in basic beliefs.[129] Thus, one way to apply the ARG criteria is to work backward from the current conclusion, through its premises, to the arguments for those premises, until we get to basic beliefs by a string of sufficient and relevant arguments.

---

[128] As Johnson & Blair (1994: p. 75) put it: "In one common sense of the term, if premises are sufficient, then you would think that's the end of the matter. However, what we mean by sufficiency is that the arguer has cited the appropriate types and amounts of evidence to support he conclusion. The relevancy and sufficiency requirements both concern her relationship of the premises to the conclusion." Other efforts to clarify this distinction are Govier's (1997: p. 74) definition of sufficiency (which we cited above) in terms of the connection between premises and conclusion; and her (p. 204) explicit statement that deductive arguments may be sufficient without having acceptable premises. Johnson & Blair's (1994: p. 269) test for sufficiency involves trying to think of a way that the premises could be acceptable and the conclusion not acceptable. One of Freeman's (1991: pp. 38-39) dialectical sufficiency questions is, "*Given* your reason, how confident should I be of your claim?" Unfortunately, there are plenty of instances where *sufficiency* is used more loosely, and seems to include both the acceptability of the premises and their link to the conclusion. This ambiguity may conceal some of the difficulties discussed here.

[129] The foundationalist argument for basic beliefs is based on the epistemological conundrum discussed earlier. There are just three possibilities: The justification process is infinitely repeatable and, thus, C is never grounded (the skeptical solution); the process circles back on itself (the coherentist solution); or it comes to an end in beliefs that do not require justification by other beliefs, either because they are arbitrary assumptions (the relativist solution) or because they are self-justifying (the foundationalist solution). Arguing in a circle is counted among the fallacies in standard informal logic treatments. That is, a conclusion cannot turn up among the premises that are used to justify itself, either in a single argument or in a series of arguments no matter how long. Since an infinite regress seems out of the question, the only solution, if beliefs are to be justified at all, is that the process come to an end. And if this process is to provide real justification, it must come to an end in beliefs that do not need to be justified, i.e., in *basic beliefs*.

The usefulness of *serial* argument structures arises more from the cognitive limitations of the reasoner than from any abstract feature of rationality. Any long derivation could in principle be shortened to a single step from the basic premises to the final conclusion. If each step in the original serial argument was deductively valid, the new single-step argument is also deductively valid. But for premises to be *sufficient*, a person must be able to *see* their probative connection with the conclusion, and this leads to a demand for shorter, relatively easy steps. To respond to this need, and break the cognitive task down into a small number of relatively simple argumentative steps, informal logic adopts the tactic of defining sufficiency independently of acceptability.

The other ARG condition, relevance, also enters the picture as a constraint that helps mitigate cognitive limitations. On the definition sufficiency, the set of premises may contain many useless elements, which have no effect on the probability of the conclusion. If argument evaluation required an assessment of the acceptability of these irrelevant premises, the best outcome would be unnecessary effort. At worst, the result might be confusion, mistaking the irrelevant for the essential and failing to supply premises that in fact are necessary for the inference. A variety of errors are classified as fallacies of irrelevance for this reason (e.g., using *ad hominem* or personal attacks to divert the course of an argument). Thus, it is desirable to require that each individual premise in an inference be relevant, and that relevance be determined independently of acceptability.[130]

There are a number of major obstacles to the foundationalist / informal logic program of defining acceptability in terms of (i) basic beliefs and (ii) a series of sufficient and relevant inference steps grounded in basic beliefs. First, there is a problem with the idea that sufficiency can be determined independently of acceptability, except in a formal logic context. Second, there is a problem with the idea of accumulating new beliefs one argument at a time. The resulting system of beliefs may be far from the best available *overall* view. Reasoning may require arguments about mental models instead of, or in addition to, arguments about individual hypotheses. Finally, the distinction between basic and non-basic beliefs does not hold up. *Any* belief, including so-called basic beliefs, may be evaluated in the light of other beliefs, and may be rejected if it does not fit into the overall best system of beliefs (Quine, 1953, 1960). By the same token, any belief may serve as part of the justification of other beliefs, even of beliefs that seem "more basic," if they play a role in the best overall mental model. We will take a look at these problems in turn.

Do We Evaluate Premises and Inferences Separately?

In non-deductive inference, unlike formal deductive logic, there can be good arguments on both sides of an issue (Johnson, 2000; Govier, 1987). The existence of conflicting arguments is part of the motivation for developing informal logic. Thus, it is surprising, to say the least, that informal logicians have devoted so little attention to how conflicting opinions should be resolved. A rather large number of textbooks address argument analysis and evaluation without discussing conflicting arguments at all. Others who do, like Thomas and Govier, recommend that conflicting positions be represented in the same diagram and that they be "balanced" against one

[130] Govier's recommendation that acceptability be considered before relevance and sufficiency would only make sense in the context of a forward chaining processes, which begins with basic premises and works toward conclusions. In a dialogue context, on the other hand, reasoning is more likely to proceed backward from a conclusion that is challenged, to reasons for that conclusion, and so on.

another. But this does not capture the dynamic aspect of verbal argumentation, in which coherent positions are developed in response to challenges from other positions, and in which conflict may be resolved by generating new alternatives. Competing positions are not generally best resolved by balancing static arguments against one another.

The standard informal logic approach to evaluation of conflicting positions has two characteristics, both of which lead to trouble:

1. Opposing positions should be combined into a single converging, balance-of-considerations argument
2. The *acceptability* of the premises and the *sufficiency* of inferential links in an argument can and should be assessed independently.

These two conditions are jointly inconsistent and, more importantly, individually incorrect. A more satisfactory approach, both normatively and empirically, is to construct coherent accounts, or mental models, for the opposing positions, and evaluate their plausibility in a way that does not distinguish between premise acceptability and sufficiency of inferential links.

Let us consider an extreme case of conflicting positions, one in which a cognizer presents a plausible, normatively adequate argument for an uncertain prediction, which then turns out to be wrong. Suppose for example that B uses the argument in Figure 37 to predict that the enemy will attack in the north (if it attacks at all) since it has no bridging equipment. An intermediate conclusion in that argument is that the enemy will *not* attack in the south, and we will consider the argument just to that point. The next day, the enemy does attack, but comes through the southern pass, contrary to B's expectations. Figure 40 shows the relevant part of B's argument from Figure 37 together with the new, surprising information that contradicts it. The new information consists of basic beliefs about testimony and observations pertaining to the actual attack. The "opposing positions" in this example are B's original argument and the basic beliefs that conflict with the prediction.

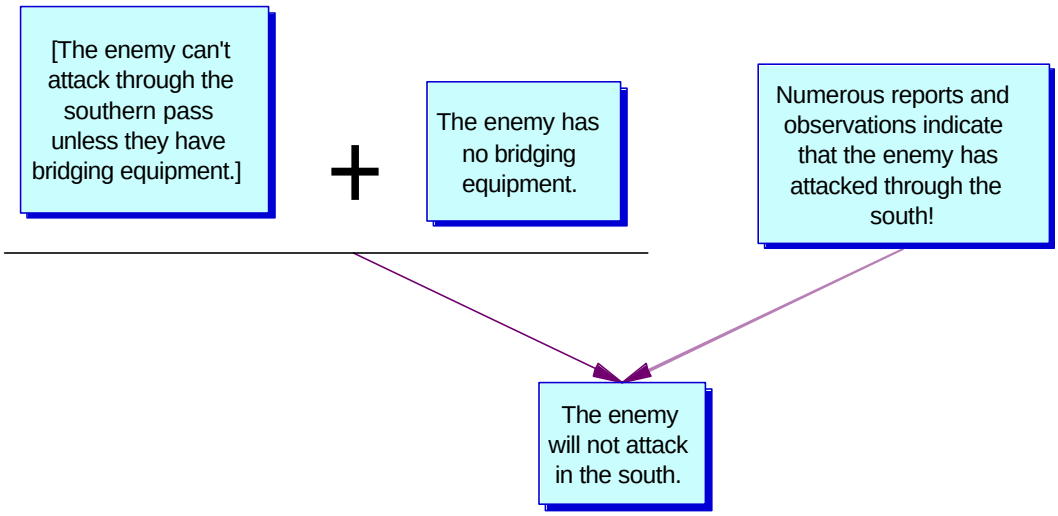


Figure 40. Information that the enemy has attacked through the south is represented by the dotted branch on the right, indicating evidence that conflicts with the conclusion.

One approach would be to keep the two arguments separate and subject them each to the ARG criteria. On the one hand, we have B's argument (Figure 37), and on the other, we have a

new argument based on the observation of the attack in the south. This will not resolve the conflict, however, because if B's argument passed the ARG criteria before (and let us suppose it did), it will do so again. According to the ARG criteria, both *attack in the south* and *not attack in the south* are supported by cogent arguments. But clearly, both conclusions cannot be accepted simultaneously.

An alternative approach, as we have seen, is suggested by Govier (1997, chapter 11; 1987), and is adopted in Figure 40. She calls attention to what she calls *conductive* inference, or *balance of consideration arguments*. In such arguments, diverse considerations are put forward, but since no single one of them is sufficient to establish the conclusion, they must be added together in a single argument in order to decide whether to accept the conclusion. When there is conflict, on the other hand, there may be sufficient reasons for more than one conclusion. But Govier treats both insufficient and multiply sufficient (conflicting) reasons as instances of the same type, and advises combining all the considerations, both pro and con, into a single converging argument. To evaluate the adequacy of such an argument, we then apply the ARG criteria to the combined argument. Unfortunately, this does not work either.

We start with B's original argument, P1 and P2, and then encounter a contrary argument involving a single premise, P3. So now we have the following combined argument:

> P1. *The enemy has no bridging equipment.*
>
> P2. *The enemy cannot attack through the south unless they have bridging equipment.*
>
> P3. *Reports and observations indicate that the enemy has attacked in the south!*
>
> C. *The enemy will not attack in the south.*

While P1 and P2 support C, P3 is a counter-consideration that strongly supports the negation of C. A decision regarding C presumably depends on the force of the case for C based on P1 and P2 versus the force of the case against C based on P3. We would expect P3 to win in this process, and to conclude that the enemy did attack in the south despite B's prediction. Unfortunately, the ARG criteria settle the issue prematurely and incorrectly.

Premises are sufficient for a conclusion if (or to the degree that) it is difficult to imagine a state of affairs in which the premises are true and the conclusion is false. In conductive, balance-of-consideration arguments, according to Govier, we evaluate sufficiency by assuming the truth of all the premises and then weighing the pros against the cons in terms of their support for the conclusion (p. 392). Unfortunately, however, the balancing process is short-circuited when one or more of the converging arguments is deductive, as in this example. For a deductive argument, sufficiency is guaranteed by deductive validity (Govier, 1997, chapter 7). If the premises deductively entail the conclusion, the conclusion cannot be false if the premises are true; thus, the argument is sufficient even if the premises are in actual fact false. According to Govier (1997: p. 203), "An argument such as this [i.e., deductively valid] is entirely adequate as far as the (R) [relevance] and (G) [good grounds, or sufficiency] conditions are concerned, so any question about its cogency must turn on the acceptability of its premises."[131]

---

[131] Actually, an argument can be deductively valid and still might contain irrelevant individual premises that are not needed in the deductive inference.

The combination of P1, P2, and P3 cannot be *sufficient* for the conclusion that the enemy attacked in the south, despite the fact that P3 strongly points in that direction. The problem is that the three premises together logically entail that the enemy will not attack in the south, by virtue of P1 and P2. The ARG criteria would not permit us to conclude that the enemy attacked in the south even if P3 were *known* to be true and P1and P2 were *known* to be false! The sufficiency criterion in this case completely eliminates the role of acceptability of premises in argument evaluation.

The *acceptability* criterion is equally problematic in this context, i.e., for conductive arguments that balance conflicting positions. Suppose that P3 is acceptable, but P1 and P2 are not. According to the ARG criteria, the argument that the enemy attacked in the south based on observations and testimony (P3) fails unless P1 and P2 are also acceptable. But this is absurd, since P1 and P2 provide a counter-consideration to the conclusion that the enemy attacked in the south. It is perverse to say that the argument is not cogent unless *all* its relevant premises are acceptable. The falsity of a counter-consideration should strengthen, not weaken the argument. Yet in judging acceptability according to the ARG method, we are not supposed to take inferential links into account. The acceptability of P1 and P2 is a separate criterion of argument cogency, in which we cannot even consider the role of P1 and P2 as counter-considerations.

Suppose we add another premise, so that *both* of the opposing arguments are deductive:

> P4. *If the enemy are observed to attack in the south, then the enemy attacked in the south.*

Now we have *two* deductive cases, P1+P2 versus P3+P4, which entail contradictory conclusions. Presumably, a decision about the conclusion should now be determined by comparing the acceptability of P1 and P2 with the acceptability of P3 and P4. Regardless of the outcome of that comparison, however, this argument cannot be cogent according to the ARG criteria because it has inconsistent premises. In formal deductive logic a proof can be constructed for any proposition whatsoever based on inconsistent premises. The argument based on P1-P4 can no more support C or not-C than it supports the claim that the moon is made of green cheese.[132] Informal logic has not improved on the limitations of formal logic in its handling logical inconsistency.

These examples happened to involve a deductive component. But the same problems arise when there is no deductive component at all. For example, consider the combined argument in Figure 39, which represents B's argument for predicting enemy attack in the north, based on lack of bridging equipment, and A's argument against the likelihood of attack in the north, based on the location of artillery. Neither argument is deductive as presented in Figure 39. But suppose B's premise, if true, would provide overwhelming support for the conclusion, while A's premise, if true, would provide merely strong support. Then, the presence of B's premise in the combined argument would make it impossible for the combined argument to provide sufficient support for A's conclusion. It doesn't matter how much more *acceptable* A's premise is in comparison to

---

[132] There is a related problem with the relevance part of the ARG method. Since inconsistent premises imply every proposition, including C, the falsity of any one premise will either reduce the probability of C or leave it unchanged. So, the premises can at best be negatively relevant to C, including P1 and P2, which supposedly support C. Any premises at all that are inconsistent with one another would be negatively relevant to any conclusion in this sense. Woods (2000) discusses some of these problems with relevance criteria.

B's premise. The very presence of B's premise blocks the cogency of any argument for A's conclusion.

The ARG method stumbles in the attempt to treat acceptability of premises and sufficiency of inferences as independent necessary conditions for cogency, and at the same time to resolve conflicting positions by integrating them into a single argument. Locating acceptability and sufficiency in separate stages of argument evaluation (i) allows arguments to be undermined merely by the possibility of a counterargument, (ii) prevents trading off strengths in acceptability against weaknesses in inferential connection or vice versa, and (iii) results in the absurd requirement that both supporting and countervailing premises must be acceptable.

All the examples so far have involved conflicting positions. But the notion that acceptability and sufficiency are independent necessary conditions is implausible even when conflicting positions are not involved. That notion depends on a hard distinction between premises and inferences that will always be, to a large degree, arbitrary outside the context of formal deductive logic. An uncertain inference can always be turned into a deductively valid one by supplying an appropriate conditional, such as P2 and P4 in the example. We argued earlier that supplying conditionals is not always a wise thing to do. (Because of open-ended defeasibility, it will be necessary occasionally to add further conditions to the antecedent of such a conditional to limit its scope. But the revised conditional will yield an inappropriate burden of proof in subsequent situations, where the original conditional should be assumed to apply by default, i.e., where defeaters are assumed false unless shown to be true.) Nevertheless, participants in argument sometimes do verbalize such conditionals, thus reframing uncertainty about the sufficiency of the inference as uncertainty about the acceptability of the conditional premise. This should not be treated as a momentous change in the nature of the argument.[133] Moreover, uncertainty can always be transferred from the inference to conditions that would, if true, defeat the inference. The falsity of a defeater might be regarded as necessary for the sufficiency of the inference, or else the negation of the defeater might be added as a premise. This is largely a distinction without a difference.[134]

In sum, the standard approach to argument evaluation in informal logic, i.e., ARG, does not give plausible results, especially in the case of conflicting opinions. It makes more sense combine acceptability of premises and sufficiency of inferences *for each of the competing positions separately*, before evaluating or comparing the positions. More generally, it takes an arbitrary aspect of how an argument happens to be represented and treats it as if it were a substantively important distinction.

Do We Evaluate Beliefs One Argument at a Time?

B was persuaded (Figure 37) by a cogent argument for what turned out to be a false conclusion, that the enemy would not attack in the south. B may wish to learn *why* his original argument went wrong in order to avoid similar errors in the future; for example, by correcting

---

[133] Our main point earlier was that *formal* deductive validity is not the distinguishing feature of *deductive argument*. It is rather the adoption of dialogical rules that temporarily restrict the types of objections (i.e., defeaters) that may be raised, so that an implication can be established in a circumscribed context. Thus, articulating a premise that makes an argument formally valid is not necessarily a sign that the dialogue is closed to defeaters, hence, deductive.

[134] We mentioned earlier that the proponent does not have a burden of proof with respect to defeaters. But this could be handled by treating certain premises (i.e., the negations of defeaters) as presumptively true unless there is reason to think otherwise.

erroneous assumptions he made about the enemy or about his sources of information. So, B now asks whether, in the light of the attack in the south, he should change his mind about any of his other beliefs, and if so, how? Here, informal logic is largely silent. From the foundationalist / informal logic point of view, the primary tool of reasoning is *serial argument* (Figure 34) which (i) focuses on an individual hypothesis as its conclusion at each step, and (ii) having once added a hypothesis to the store of accepted beliefs does not circle back to reconsider the way the hypothesis was arrived at. The addition of *defeasibility* in contemporary foundationalism allows the retraction of a conclusion that turns out to be wrong (Figure 36), as well as the retraction of premises or inferences by means of which the conclusion was derived. However, when more than one premise or inference is defeasible, no strategy or even guidance is offered within either informal logic or contemporary foundationalism for discovering the specific premises or inferences that are responsible for the mistake.

In trying to learn from his mistaken prediction, the first problem that B faces is that very little of his reasoning was explicit in Figure 37. The failure of the prediction might therefore prompt him to bring more of his reasoning to the light of day. For example, he believed that the enemy would need bridging equipment in order to attack in the south. Further, suppose he now recalls that it was a report from the intelligence (G-2) staff that led him to believe that the enemy had no bridging equipment. That report supported the conclusion of the original argument because he assumed it was reliable. So B now considers a defeater to the original inference: The G-2 report implies that the enemy has no bridging equipment, *unless the report is unreliable*. Adding these elements to his original argument, as shown in Figure 41, gives him more to work with in identifying the source of error.

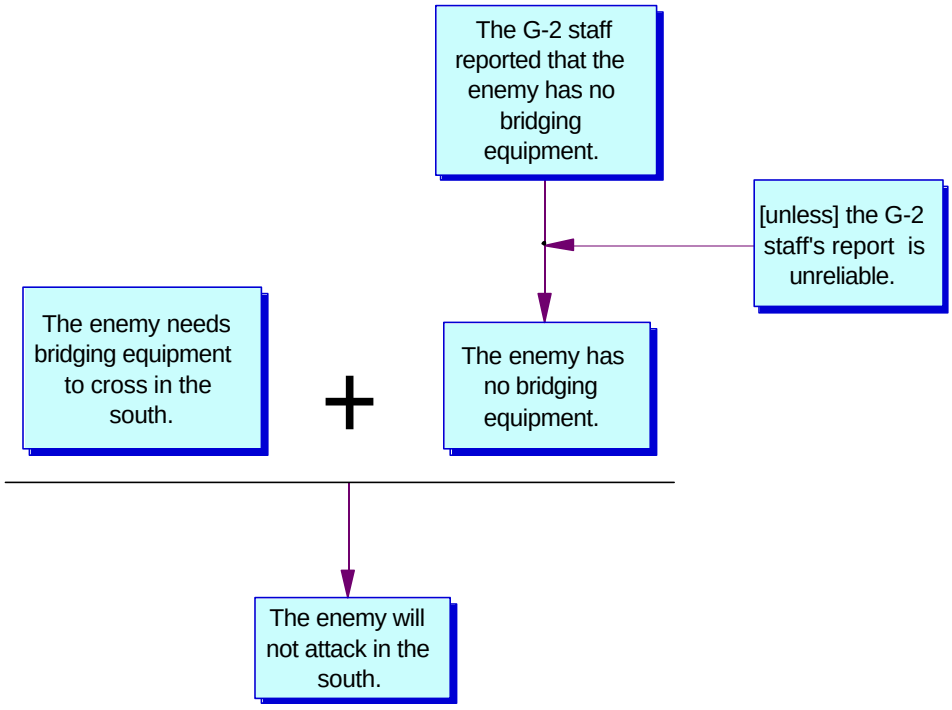


Figure 41. B's original argument that the enemy would not attack in the south, after being expanded to make other premises and inferences explicit.

B can now use the failure of the prediction about attack in the south to create a *new* argument that the defeater in Figure 41 is in fact the case. Figure 42 shows how the surprising

attack in the south plus the premises of the argument in Figure 41, provide a straightforward argument that the G-2 staff's report was unreliable. If the enemy attacked in the south, they must have had bridging equipment; but the G-2 staff reported that they did not; hence, the G-2 report was false. If the original argument was cogent according to ARG criteria, and if there is no strong independent evidence for the reliability of the G-2's report, then the new argument in Figure 42 will also be cogent. Thus, B is entitled to conclude that the G-2 report was unreliable.

But suppose B's colleague C, who agreed with B's original prediction about no attack in the south, is now thinking about this a different way. He asks himself why they believed bridging equipment was necessary in the first place, and recalls an engineering staff report about the fordability of rivers in the south. That report supported the conclusion of the argument only because they believed it was reliable. So C adds another defeater to the argument: The engineers' report about the fordability of rivers indicated that bridging equipment was necessary *only if it was reliable*. The result is a further expansion of the original argument, as shown in Figure 43. C can now use the failure of the original prediction to create a new argument that *this* defeater is in fact the case. Figure 44 shows C's argument that the engineering staff's report was unreliable. If the original argument was cogent according to ARG criteria, and if there is no strong independent evidence for the reliability of the engineering report, then this new argument is also cogent. Thus, C is entitled to conclude that the engineering staff report was unreliable.

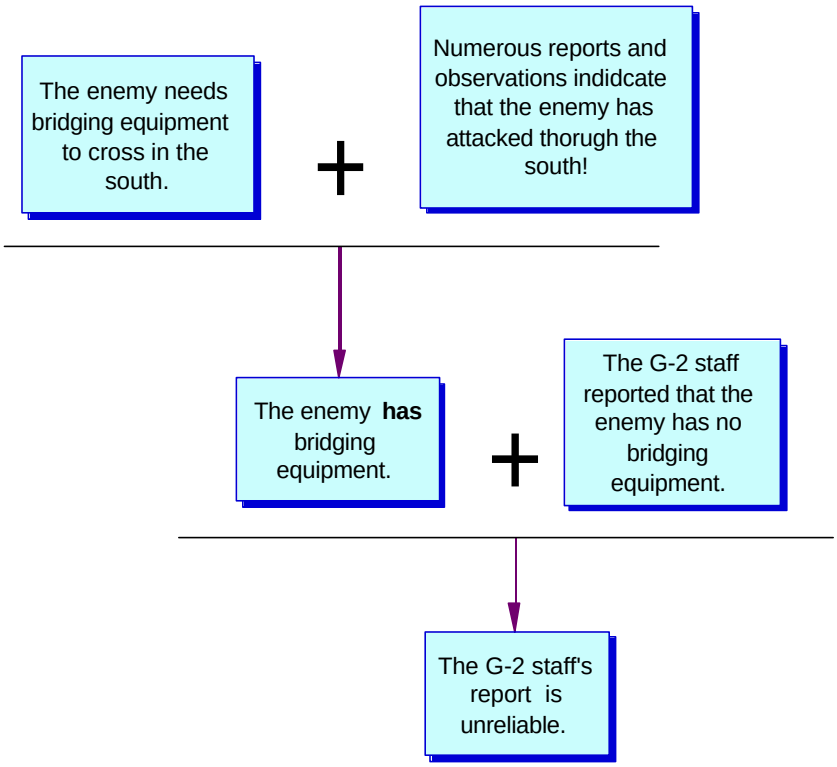


Figure 42. B's new argument that the G-2 staff's report was unreliable.

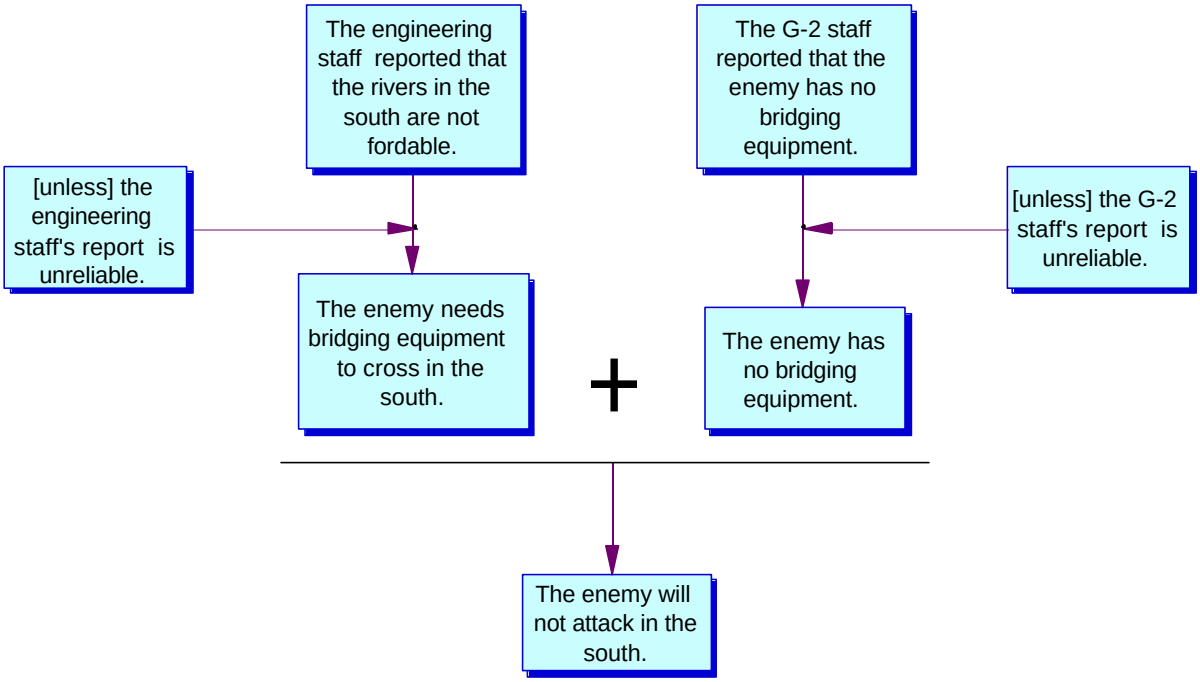


Figure 43. C's further expansion of B's original argument, to make additional premises and inferences explicit.

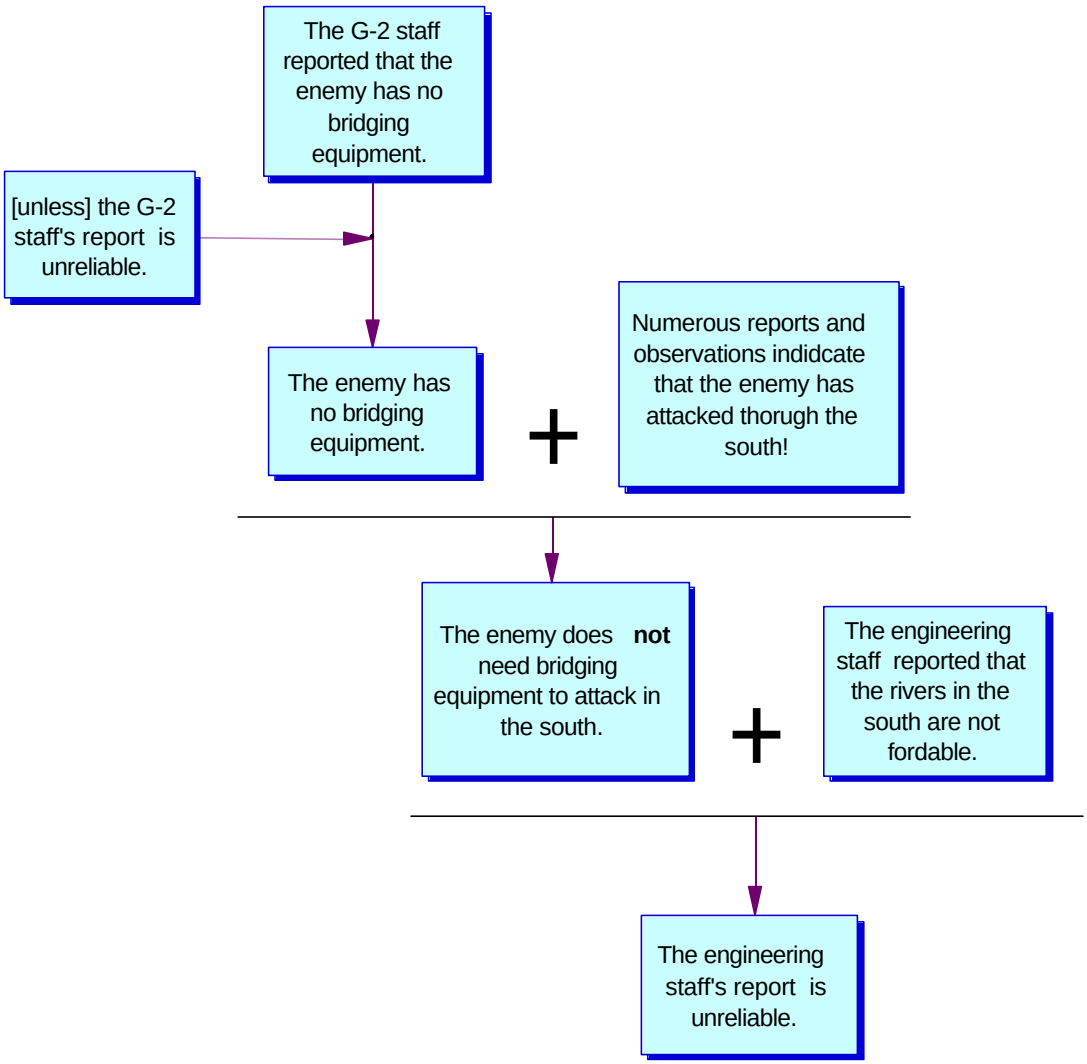


Figure 44. C's new argument that the engineering staff report was not reliable.

According to the argument-centered strategy promoted by informal logic, B and C can create separate arguments in favor of possible revisions of individual beliefs, and then evaluate each one for cogency. Thus, Figure 42 and Figure 44 pinpoint different parts of the original argument as culpable for the mistaken prediction. And of course it is possible to expand the argument still further to identify additional potential culprits. Unfortunately, this can and does lead to unsatisfactory results. First, it is easily possible for *all* such arguments (like Figure 42 and Figure 44) to be cogent according the ARG criteria. This would lead a cognizer to make more revisions in his current beliefs than is necessary to resolve the conflict, e.g., to conclude that *both* the G-2 *and* the engineering report was unreliable. This is almost certainly a less plausible verdict than that only one of them was mistaken. The *combination* of all the original premises and inferences led us into trouble. But it is not necessary to *drop* all of them in order to achieve a coherent explanation. In fact, since each was considered plausible in the first place, the simultaneous falsity of all of them might be extremely improbable.

Under other circumstances, it is possible that *none* of the arguments for belief revision will be cogent. This would happen, for example, if B and C have independent reasons to believe

in the reliability of the G-2 report and engineering report respectively. Such reasons would be included in Figure 42 and Figure 44 as conflicting arguments, and could render the premises insufficient to establish the conclusion in each of those arguments.[135] But this outcome is also unsatisfactory. Even though each of the original beliefs and inferences is still more likely than not, and all the arguments against them are insufficient, nevertheless, *all* of these beliefs and inferences cannot be correct. At least one of them must be revised in order to arrive at a coherent overall set of beliefs. Yet if the ARG criteria allow no additional belief revisions even after the surprise attack in the south, the cognizer will be left with an incoherent set of beliefs.

Do We Evaluate Mental Models?

Revision of incoherent beliefs is not effectively handled by the one-hypothesis-at-a-time, argument-by-argument method. Effective belief revision cannot be accomplished by a series of separate choices that, claim by claim, decides between each claim and its negation. The result is not the best overall account. Everitt and Fisher (1995: p. 173) share our characterization of standard approaches in informal logic: "The prevailing orthodoxy is that it *is* possible to consider at least some arguments in isolation from anything else and determine whether they are good or not." The solution is to frame the problem differently. The arguer wants to end up not with a cogent argument, but with a *single coherent account* of the situation, i.e., the most acceptable mental model.

Table 11 will illustrate how a mental model-based approach might apply to this simple example, in the absence of any capacity constraints. Each of the numbered rows in Table 11 is a mental model in our example. The dimensions of variation among the models are the five elementary propositions (the top column in the table), each of which may be true or false. These include three basic beliefs, of which B and C are very confident: that the G-2 staff made its report, that the engineers made their report, and that the enemy attacked in the south. In addition, the dimensions include two inferences of which B and C are less confident, which are represented by the negation of the defeaters in Figure 43: i.e., that the G-2 staff is not unreliable and that the engineering staff is not unreliable. (If the negations of the defeaters associated with those inferences are true, the inferences are valid.)

For concreteness, we assign independent probabilities of .99 each to the three basic beliefs. We will also imagine that B has prior reason to trust both the G-2 staff and the engineering staff, but that he trusts the G-2 staff slightly less on these topics. The inference regarding the G-2 thus gets a probability of .86, and the inference regarding the engineers gets a probability of .90. We know that all five of the elementary propositions cannot be true, since they lead to a contradiction (*the enemy attacks in the south* and *the enemy does not attack in the south*). Assuming that B and C choose not to question the basic premises, that leaves three consistent combinations of truth and falsity of the two defeaters. The far right column of Table 11 shows their probabilities.[136]

---

[135] This is not the same problem that we discussed in the last section. There, we pointed out the inability to trade off acceptability and sufficiency. Here, we have separate arguments directed against possible culprits in a failed prediction. All the premises of each argument might be *acceptable*, but all the inferences *insufficient* (because of conflict with prior evidence supporting the reliability of the culprit under examination). Thus, we are not justified in rejecting any of the suspects, even though we know that at least one of them must be guilty and the conclusion of at least one of these arguments must be true.

[136] The probability of a mental model is proportional to the probabilities of its components. The probabilities in the

Table 11. Numbered rows are alternative mental models for the situation in Figure 40. Each mental model is a different combination of truth (T) and falsity (F) of the propositions in the top row.

	G-2 reports no bridging equipment. (.99)	G-2 is not unreliable. (.86)	Engineers report unfordable rivers. (.99)	Engineers are not unreliable. (.90)	Enemy attacks in the south/ (.99)	**Probability of mental model**
1	T	T	T	T	T	.00
2	T	T	T	F	T	.34
3	T	F	T	T	T	.49
4	T	F	T	F	T	.06

Table 11 enables B and C to compare complete alternative situation pictures. The combination in which *both* the G-2 and the engineers are unreliable is unlikely, with a probability of .06. The two combinations in which only *one* of them is unreliable are more likely. And the most likely model of the three is the one in which the G-2's report is unreliable and the engineering report is reliable (with a probability of .49). This mental model might lead B to scrutinize G-2 reports on this topic more carefully in the future.

There are some important morals of this example. When there are conflicting opinions, a cognizer must never regard a single argument as the last word, *even if it includes all the available information* (as Figure 44 does) and *passes all the ARG criteria*. The two arguments illustrated above (Figure 42 and Figure 44) cannot solve the problem either individually or jointly. If the cognizer considers only one of them, she runs the risk of dropping a belief that should be kept, or of retaining a belief should be dropped. If she considers both of the arguments, she may either continue to hold an incoherent set of beliefs or adopt an overall view that is implausible (by revising more of her beliefs than is necessary). Moreover the two arguments (Figure 42 and Figure 44) cannot be diagrammed as parts of a single converging argument, as suggested by Thomas and Govier for conflicting arguments, since they do not pertain to the same hypothesis.

Ultimately, the problem with arguments for individual hypotheses is due to *defeasibility* of inferences, i.e., the possibility of encountering new information that forces the retraction of previous conclusions. But more precisely, it is due to the symmetrical roles that alternative explanations play as defeaters for one another. Thus, Figure 44 shows that the engineering staff's report is unreliable *unless the G-2 staff's report was unreliable*. And Figure 42, if suitably expanded, would show that the G-2 staff's report is unreliable *unless the engineering staff's*

---

far right cell of each row are obtained by taking the given probability for each component if there is a T, and one minus that probability if there is an F, multiplying these together across the row, and dividing by one minus the chance that all five elements are true. The latter normalization eliminates the impossible case in which all of B's beliefs are true. The table does not include a row for every possible model, only for combinations in which the three basic beliefs are true. There are 28 additional mental models in which one or more basic beliefs are false. The total probability of the omitted possibilities is .12. Thus, the sum of the probabilities shown in the table is.88.

report was unreliable. We could have expanded this example so that B and C explored additional possible explanations of the failed predictions. If we had, each of the explanations would be a defeater for all the others. Because of this symmetry, adding defeaters to the architecture of arguments does not make arguments for individual hypotheses more useful in the resolution of conflicts. A dialogue of warring arguments will go on tit-for-tat but will not resolve conflict unless the participants are able to assemble the implications of the arguments into a larger picture.

Arguments in informal logic lead to acceptance or rejection of individual claims. No method is provided for evaluating *as a whole* the set of beliefs that results from a series of arguments, or for revising earlier conclusions based on later ones. The ARG method assumes that locally optimal decisions with respect to each intermediate conclusion will yield a globally optimal system of beliefs, i.e., an adequate overall picture of the situation. But this is not the case when different arguments point to different conclusions. Failure to consider an ensemble of interrelated beliefs as a whole can lead to impossible or implausible models of the situation, and thus to a complacency that is incompatible with the goals of critical thinking. (Similar problems arise in picking a stock portfolio stock by stock rather than considering how they relate to one another to affect overall performance.) [137]

The role of arguments is in part to probe for problems in mental models, such as incoherence. It was the conflict between the conclusions of two arguments (Figure 40) that first told B that his beliefs about this situation were flawed. But arguments for individual hypotheses cannot generally *resolve* differences of opinion. If the cognizer does rely on such arguments, she should construct a separate one for rejecting each of the alleged culprits (e.g., Figure 42 and Figure 44), compare the force of those arguments in a way that aggregates premise acceptability and inference strength, and then use the results to build a coherent overall account that involves as few changes as possible from her original view. But this is equivalent to selecting the mental model with the highest probability. In this example that strategy can be implemented by revising the weakest element in the original model (i.e., the belief in the G-2's reliability).

In our example, one of the conflicting "opinions" was virtually certain knowledge that the enemy had attacked in the south. Thus, B's motivation for comparing alternative possibilities in Table 11 was to find and correct erroneous assumptions in the body of beliefs that led to the incorrect prediction. However useful this might be, our point is more general: Learning from mistakes is not the only reason for comparing mental models. It is also necessary in order to settle the conflict itself when both sides are uncertain. For example, before the attack took place, B predicted attack in the north based on the absence of bridging equipment, while A predicted attack in the south based on the location of enemy artillery. Both A and B have an argument supporting their respective positions (Figure 37and Figure 38). The dispute can be resolved only by going beyond these arguments to situation pictures that explain *why* the opposing argument might fail. For example, in order to make the case that the enemy will attack in the south, A may argue not only that this fits the artillery location, but that a mental model in which the G-2 was

---

[137] Pearl (1989) makes a similar point in the context of a probabilistic framework: "…by belief commitment we mean the categorical but tentative acceptance of a subset of hypotheses that together constitute the most satisfactory explanation of the evidence at hand. In probabilistic terms, that task amounts to finding the most probable instantiation of all hypothesis variables, given the observed data.[p. 240] …this optimal assignment cannot be obtained simply by optimizing the belief distributions of the individual variables [p. 246] ."

unreliable is more likely than one in which the enemy plans to attack without artillery preparation. B may respond by creating a mental model in which his view is true and A's is false, e.g., in which the G-2 and engineers are reliable, but the enemy has longer range artillery. Conflict resolution involves a dialogue in which each side probes for weaknesses in the other side's case. The purpose of that probing is to build plausible mental models that explain how the other side's case might be false. Conflict resolution itself requires that the parties either explicitly or implicitly compare the resulting mental models.

Why Accept the Conclusion of a Long Argument?

There is a related, but still more general problem that is solved when we switch focus from arguments about individual hypotheses to arguments about alternative mental models. Informal logicians define non-deductive inference as reasoning that makes its conclusions *probable* rather than *certain*. Unfortunately, the ARG criteria do not accomplish even that. They often dictate the acceptance of extremely low probability conclusions. The problem arises because the acceptability of the conclusion of an argument depends on the likelihood that *every* essential component of the reasoning is true. The probability of the conclusion decreases exponentially with the number of components that are necessary to the argument. Thus, if we extend the chain of reasoning or require linked premises at each step, the chance of a conclusion's being true can be made arbitrarily low (unless offset by a comparable increase in converging arguments for the same conclusion). If there happens to be a conflicting argument, then the failure of that argument is also necessary, through the falsity of a premise or inference in the conflicting argument; this reduces the probability of the conclusion still further. But even without conflict, a belief that is supported by an even moderately complex argument will acquire little credibility at the end. Nevertheless, the conclusion may be accepted according to ARG criteria because *each step* in the serial argument, taken separately, is cogent. Thus, the ARG criteria, applied in a step by step fashion reminiscent of formal deductive logic, are in conflict with their own probabilistic rationale. It may seem reasonable to accept each intermediate conclusion, but in the end one is out on a long limb.

A similar lesson follows from an example called the *lottery paradox*. If you buy a ticket in a lottery for which a million tickets have been sold, the chance of your winning is extremely low, and the chance of loosing is correspondingly high. You are therefore justified, according to *this* argument, in concluding that you will not win. (If the probability of loosing is not high enough, you can always imagine more tickets have been sold. Consider also that many of the beliefs you do accept and act on every day are much less probable than this one!) But the *same* argument could be made for each of the million tickets, implying that no one will win. Since you know that someone will win, the result of accepting the conclusion of each individual argument is that you have incoherent beliefs. Moreover, it makes no sense to resolve this inconsistency by randomly selecting a ticket and concluding that *it* will win. An overall view of the matter is necessary, and it suggests suspending judgment.

A better interpretation is to view reasoning not as accepting beliefs one by one, but as evaluating and accepting mental models. The *options* are complete models as opposed to individual hypotheses. The cognizer is attempting to put together a workable picture of the whole situation (more precisely, a picture of the parts that are relevant to her), and that picture includes premises, negations of defeaters, truth of defeaters for conflicting arguments, as well as other hypotheses of interest (the "conclusions"). She cannot require a complete model to have an absolute probability greater than .50, or indeed, greater than any particular threshold. The

absolute probability of a model depends on the number of components of the situation that are relevant enough to be included in the model, and thus the absolute probability has no rational bearing on the acceptability of the model as a whole. Comparison to equally complete models is all that counts.[138]

The need for lengthy serial arguments originates from the doctrine that some beliefs are basic, hence suitable for grounding inferences, and others are not. According to the contemporary foundationalist paradigm, only basic beliefs have prior credibility that an argument can build on. Thus, a cognizer must trace every conclusion back to basic beliefs by some chain of reasoning, no matter how many steps it contains. But as we move forward from basic beliefs along a chain of cogent inferences, the probability of each newly justified belief is attenuated by the additional inference step used to reach it and by each linked premise that is needed for that step. Thus, the requirement for grounding by basic beliefs helps generate the problems we have just discussed.

Contemporary foundationalism views the distinction between *basic* and *non-basic* beliefs as a difference in kind. Indeed, Table 11 only looks at combinations in which the three so-called basic beliefs are true. But this was a *choice* that B and C made owing to the small chance that these particular beliefs were false, compared to the other beliefs and inferences. Even though the probabilities are high, it is possible for a so-called basic belief to be false. They are the last beliefs to be called into question, but are not immune from doubt altogether. It is not difficult to imagine situations in which multiple lines of reasoning that are inconsistent with a basic belief would lead to its revision. (For example, suppose B misunderstood the G-2's report; the part of the report he saw was about another enemy unit. Thus, the G-2 did not actually report that the relevant enemy unit had no bridging equipment.) The most probable overall account may on occasion involve abandoning a basic belief. This insight spells trouble for contemporary foundationalism, which Haack classifies as "weak" and "impure" in comparison to the classical version:

> Weak foundationalism concedes that basic beliefs need not be fully justified by experience alone; but then what reason remains to deny that they could get more (or less) justified by virtue of their relations to other beliefs? Impure foundationalism concedes that there can be mutual support among derived beliefs…but now the insistence that derived beliefs can give no support to basic beliefs looks arbitrary, and the distinction of basic and derived beliefs pointless. (Haack, 2000/1999: p. 227)

Arguments do not ground non-basic beliefs in basic beliefs. Rather, they make explicit clusters of beliefs that are particularly relevant to a situation, out of a background of many other

---

[138] Another way has been proposed for responding to the problem of low probability conclusions at the end of long serial arguments. This is to abandon the standard rules for manipulating probabilities. For example, Pollock (1995: pp. 95-99) proposes that for epistemic probabilities, the probability of a conjunction is not the product of the component probabilities. Rather, it is the weakest link, i.e., the *minimum* of the various probabilities that would ordinarily be multiplied together. In that case, the probability of the final conclusion of a large argument would be the probability of the least probable component in the argument for that conclusion. More generally, the probability of a conclusion would always exceed a threshold when every step in the argument exceeded the threshold. This solution is ad hoc and designed to salvage the step-by-step approach to justifying a conclusion. It is unnecessary if one adopts the position that complete alternatives are the options that one is evaluating.

beliefs. The ultimate justification of that cluster of beliefs lies in their relationships not only to one another but to the implicit background.

REFERENCES

Alfaro-LeFebre, R. (1999). Critical thinking in nursing: A practical approach. Philadelphia, PA: W.B. Saunders.

Allston, W. (2000). How to think about reliability. In E. Sosa & J. Kim (Eds.), Epistemology: An anthology. Malden MA: Blackwell Publishers.

Anderson, J. (1982). Acquisition of cognitive skill. Psychological Review, 89(4), 369-406.

Anderson, J.R. (1990). The adaptive character of thought. Mahwah, NJ: Lawrence Erlbaum Associated, Inc.

Anderson, T., Howe, C., & Tolmie, A. (1996). Interaction and mental models of physics phenomena: Evidence from dialogues between learners. In J. Oakhill & A. Garnham (Eds.), Mental models in cognitive science (pp. 247-274). East Sussex, UK: Psychology Press.

Astington, J.W., Harris, P.L., & Olson, D.R. (Eds.) (1988). Developing theories of mind. Cambridge MA: Cambridge University Press.

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

Audi, R. (1998). Epistemology. NY: Routledge.

Baron, J., & Brown, R.V. (Eds.) (1991). Teaching decision making to adolescents. Mahwah NJ: Lawrence Erlbaum Associates.

Barth, E.M. (1987). Logic to some purpose: Theses against the deductive-nomologic paradigm in the science of logic. In F.H. van Eemeren, R. Grootendorst et al. (Eds.), Argumentation: Across the lines of discipline. Dordrecht-Holland: Foris Publications.

Beardsley, M.C. (1950). Practical logic. New York: Prentice Hall.

Bell, D.E., Raiffa, H., & Tversky, A. (Eds.) (1998). Decision making: Descriptive, normative, and prescriptive interactions. Cambridge: Cambridge University Press.

Berry, D.C., & Dienes, Z. (1993). Implicit learning. Mahwah NJ: Lawrence Erlbaum.

Bogdan, R.J. (2000). Minding minds. Cambridge MA: MIT Press.

Bonjour, L. (1985). The structure of empirical knowledge. Cambridge, MA: Harvard University Press.

Bonjour, L. (2000). Can empirical knowledge have a foundation? In E. Sosa & J. Kim (Eds.), Epistemology: An anthology Oxford. UK: Blackwell.

Bonjour, L. (2001). Toward a defense of empirical foundationalism. In M.R. Depaul (Ed.), Resurrecting old-fashioned foundationalism. Boston MA: Rowman & Littlefield.

Brandom, R. (2000). Knowledge and the social articulation of the space of reasons. In E. Sosa & J. Kim (Eds.), Epistemology: An anthology. Malden MA: Blackwell Publishers.

Brookfield, S.D. (1987). Developing critical thinkers. San Francisco: Jossey-Bass.

Browne, M.N., & Keeley, S.M. (1994). Asking the right questions: A guide to critical thinking. Upper Saddle River, NJ: Prentice-Hall.

Browne, M.N., & Keeley, S.M. (1998). Asking the right questions: A guide to critical thinking (Fifth Edition). Upper Saddle River NJ: Prentice-Hall.

Cherniak, C. (1986). Minimal rationality. Cambridge, MA: MIT Press.

Chipman, S. F., Segal, J. W., & Glaser, R. (Eds.) (1985). Thinking and learning skills: Research and open questions. Mahwah NJ: Lawrence Erlbaum Associates.

Chisholm, R. (1977). Theory of knowledge. NY: Prentice-Hall.

Chisholm, R.M. (2000). The indispensability of internal justification. In S. Bernecker & F. Dretske (Eds.), Knowledge: Readings in contemporary epistemology. Oxford: Oxford University Press.

Cohen, M.S. (1993a). Three paradigms for viewing decision biases. In G.A. Klein, J.R. Orasanu, R. Calderwood, & C.E. Zsambok (Eds.), Decision making in action: Models and methods Norwood, NJ: Ablex Publishing Corporation.

Cohen, M.S. (1993b). The naturalistic basis of decision biases. In G.A. Klein, J.R. Orasanu, R. Calderwood, & C.E. Zsambok (Eds.), Decision making in action: Models and methods Norwood, NJ: Ablex Publishing Corporation.

Cohen, M.S. (1993c). The bottom line: Naturalistic decision aiding. In G.A. Klein, J.R. Orasanu, R. Calderwood, & C.E. Zsambok (Ed.), Decision making in action: Models and methods. Norwood, NJ: Ablex Publishing Corporation.

Cohen, M.S., & Thompson, B.B. (2001). Training teams to take initiative: Critical thinking in novel situations. In E. Salas (Ed.), Advances in cognitive engineering and human performance research. JAI.

Cohen, M., Freeman, J.T., & Thompson, B.B. (1998). Critical thinking skills in tactical decision making: A model and a training method. In J.C. Salas (Ed.), Decision-making under stress: Implications for training and simulation. Washington, DC: American Psychological Association Publications.

Cohen, M.S., Thompson, B.B., Adelman, L., Bresnick, T.A., Shastri, L., & Riedel, S. (2000a). Training critical thinking for the battlefield: Volume I: Basis in cognitive theory and research. Arlington, VA: Cognitive Technologies, Inc.

Cohen, M.S., Thompson, B.B., Adelman, L., Bresnick, T.A., Shastri, L., & Riedel, S. (2000b). Training critical thinking for the battlefield. Volume II: Training system and evaluation. Arlington, VA: Cognitive Technologies, Inc.

Cohen, M.S., Thompson, B.B., Adelman, L., Bresnick, T.A., Shastri, L. & Riedel, S. (2000c). Training critical thinking for the battlefield: Volume III: Modeling and simulation of battlefield critical thinking. Arlington, VA: Cognitive Technologies, Inc.

Conee, E., & Feldman, R. (2000). The generality problem for reliabilism. In E. Sosa & J. Kim (Eds.), Epistemology: An anthology. Oxford, UK: Propwell.

Curd, M. & Cover, J.A. (1998). The Duhem-Quine thesis and underdetermination: Commentary. In M. Curd & J.A. Cover (Eds.), Philosophy of science: The central issues (pp. 354-408). NY: W.W. Norton & Company.

Dancy, J., & Sosa, E. (1992). A companion to epistemology. In Propwell companions to philosophy. Oxford, UK: Propwell Publishers.

Dauer, F.W. (1989). Critical thinking: An introduction to reasoning. NY: Barnes & Noble Books.

Davidson, D. (1986). A coherence theory of truth and knowledge. In E. LePore (Ed.), Truth and interpretation: Perspectives on the philosophy of Donald Davidson. Oxford, UK: Propwell.

Davidson, B.W. (1998). A case for critical thinking in the English language classroom. TESOL Quarterly, Vol. 32, No. 1, Spring.

Dawes, R.M. (2001). Everyday irrationality. Boulder CO: Westview Press.

Day, T.J. (1989). Circularity, non-linear justification, and holistic coherentism. In J.W. Bender (Ed.), The current state of the coherence theory. Dordecht-Holland: Kluwer Academic Publishers.

Dennett, D.C. (1978). Brainstorms: Philosophical essays on mind and psychology. Montgomery, VT: Bradford Books.

DeRose, K. (2000). Solving the skeptical problem. In E. Sosa & J. Kim (Eds.), Epistemology: An anthology. Oxford UK: Blackwell.

Dewey, J. (1991). How we think. Buffalo, NY: Prometheus Books.

Diestler, S. (1998). Becoming a critical thinker: A user friendly manual. Upper Saddle River, NJ: Prentice Hall.

Dretske, F.I. (1983). Knowledge and the flow of information. Cambridge MA: MIT Press.

Duffy, L. (1993). Team decision-making biases: An information-processing perspective. In G.A. Klein, J.R. Orasanu, R. Calderwood, & C.E. Zsambok (Eds.), Decision making in action: Models and methods (pp. 346-359). Norwood, NJ: Ablex Publishing Corporation.

Duhem, P. (1998). Physical theory and experiment. In M. Curd & J.A. Cover (Eds.), Philosophy of science: The central issues (pp. 257-279). NY: W.W. Norton & Company.

Empiricus, S. (1990). Outlines of pyrrhonism. NY: Prometheus Books.

Endsley, M. (1997). The role of situation awareness in naturalistic decision making. In C.E. Zsambok & G. Klein (Eds.), Naturalistic decision making. Mahwah, NJ: Erlbaum.

Ennis, R.H. (1962). A concept of critical thinking. Harvard Educational Review, 32, 81-111.

Ennis, R.H. (1987). A taxonomy of critical thinking dispositions and abilities. In J.B Baron & R.J. Sternberg (Eds.), Teaching critical thinking skills. NY: W.H. Freeman.

Ennis, R.H. (1996). Critical thinking dispositions: Their nature and assessment. Informal Logic, 18(2), 165-182.

Epstein, R.L. (1999). Critical thinking. Belmont CA: Wadsworth Publishing.

Ericsson, K. (1996). The acquisition of expert performance: An introduction to some of the issues. In K. Ericsson (Ed.), The road to excellence: The acquisition of expert performance in the arts and sciences, sports and games. Mahwah, NJ: Lawrence Erlbaum Associates.

Ericsson, K.A., & Smith, J. (1991). Prospects and limits of the empirical study of expertise: an introduction. In K. A. Ericsson & J. Smith (Eds.), Toward a general theory of expertise. Cambridge UK: Cambridge University Press.

Everitt, N., & Fisher, A. (1995). Modern epistemology: A new introduction. Cambridge, MA: McGraw-Hill.

Facione, N.E. (1990). The Delphi report. Committee on pre-college philosophy. American Philosophical Association.

Feldman, R., & Conee, E. (2000). Evidentialism. In E. Sosa & J. Kim (Eds.), Epistemology: An anthology. Oxford UK: Blackwell.

Feynman, R. (1965). The character of physical law. Cambridge MA: The MIT Press.

Fisher, A. (2000). Informal logic and its implications for philosophy. Informal Logic, 20, 109-115.

Fisher, A. & Scriven, M. (1997). Critical thinking: Its definition and assessment. Centre for Research in Critical Thinking. Norwich/Edgepress, California, USA.

Fisher, S.C. & Spiker, V.A. (2000). Application of a theory of critical thinking to Army command and control. Anacapa Sciences, Inc.

Flew, A. (1998). How to think straight: An introduction to critical reasoning. Amherst, NY: Prometheus Books.

Foley, R. (2000). Skepticism and rationality. In E. Sosa & J. Kim (Eds.) Epistemology: An anthology. Oxford, UK: Propwell.

Forster, M. (1999a). How do simple rules "fit to reality' in a complex world. Minds and Machines, 9, 543-564. http://philosophy.wisc.edu/forster/default.htm

Forster, M. (1999b). Model selection in science: The problem of language invariance. British Journal for the Philosophy of Science, 50, 83-102. http://philosophy.wisc.edu/forster/default.htm

Forster, M. (2001). The new science of simplicity. In A. Zellner, A. Keuzenkamp, & M. McAleer (Eds.), Simplicity, inference, and modelling (pp. 83-119). Cambridge: Cambridge University Press. http://philosophy.wisc.edu/forster/default.htm

Forster, M., & Sober, E. (1994). How to tell when simpler, more unified, or less ad hoc theories will provide more accurate predictions. British Journal for the Philosophy of Science, 45, 1-35. http://philosophy.wisc.edu/forster/default.htm

Freely, A.J.S., & Steinberg, D.L. (2000). Argumentation and debate: Critical thinking for reasoned decision making. Belmont, CA: Wadsworth.

Freeman, J.B. (1991). Dialectics and the macrostructure of arguments: A theory of argument structure. Dordrecht-Holland: Foris Publications.

Freeman, J.B. (1993). Thinking logically: Basic concepts for reasoning. Englewood Cliffs, NJ: Prentice Hall.

Freeman, J.B. (2000). The place of informal logic in philosophy. Informal Logic, 20(3), 117-128.

Fumerton, R. (2001). Classical foundationalism. In M.R. Depaul (Ed.), Resurrecting old-fashioned foundationalism. Boston MA: Rowman & Littlefield.

Gardenfors, P. (Ed.) (1992). Belief revision. Cambridge UK: Cambridge University Press.

Gell-Mann, M. (1994). The quark and the jaguar. NY: W.H. Freeman and Company.

Gigerenzer, G., & Selten, R. (Eds.) (2001). Bounded rationality: The adaptive toolbox. Cambridge MA: MIT Press.

Gillies, D. (1998). The Duhem thesis and the Quine thesis. In M. Curd & J.A. Cover (Eds.), Philosophy of science: The central issues (pp. 302-319). NY: W.W. Norton & Company.

Glaser, E.M. (1941). An experiment in the development of critical thinking. NY: Teachers College of Columbia University.

Glymour, C. (1980). Theory and evidence. Princeton NJ: Princeton University Press.

Goldman, A.I. (1979/1992). What is justified belief? In A.I. Goldman (Ed.), Liaisons: Philosophy meets the cognitive and social sciences. Cambridge MA: MIT Press.

Goldman, A.I. (1986). Epistemology and cognition. Cambridge, MA: Harvard University Press.

Goldman, A.I. (1989/1992). Epistemic folkways and scientific psychology. In A.I. Goldman (Ed.), Liaisons: Philosophy meets the cognitive and social sciences. Cambridge MA: MIT Press.

Goldman, A.I. (1992). Liaisons: Philosophy meets the cognitive and social sciences. Cambridge, MA: MIT Press.

Goldman, A. (2001). Internalism exposed. In M. Steup (Ed.), Knowledge, truth, and duty: Essays on epistemic justification, responsibility, and virtue. Oxford UK: Oxford University Press.

Goodman, N. (1965). Fact, fiction, and forecast. Indianapolis: The Bobbs-Merrill Company.

Govier, T. (1987). Problems in argument analysis and evaluation. Dordrecht-Holland: Foris Publications.

Govier, T. (1997). A practical study of argument. Belmont, CA: Wadsworth Publishing.

Green, D. (1996). Models, arguments, and decisions. In J. Oakhill & A. Garnham (Eds.), Mental models in cognitive science. East Sussex, UK: Psychology Press.

Grice, P. (1989). Studies in the way of words. Cambridge, MA: Harvard University Press.

Groarke, L.A., Tindale, C.W., & Fisher, L. (1996). Good reasoning matters! A constructive approach to critical thinking. NY: Oxford University Press.

Guzzo, R.A. (1995). Introduction: At the intersection of team effectiveness and decision making. In R.A. Guzzo, E. Salas, & Associates (Eds.), Team effectiveness and decision making in organizations (pp 1-8). San Francisco, CA: Jossey-Bass Inc., Publishers.

Haack, S. (1993). Evidence and inquiry: Towards reconstruction in epistemology. Oxford, UK: Propwell.

Haack, S. (2000). A foundherentist theory of empirical justification. In E. Sosa & J. Kim (Eds.), Epistemology: An anthology. Malden MA: Blackwell Publishers.

Halpern, D.F. (1996). Thought and knowledge. Mahwah, NJ: Lawrence Erlbaum Associates.

Hamblin, C.H. (1970). Fallacies. Newport News VA: Vale Press.

Harman, G. (1973). Thought. Princeton NJ: Princeton University Press.

Harman, G. (1986). Change in view. Cambridge, MA: The MIT Press.

Harman, G. (1999a). Reasoning, meaning and mind. Oxford UK: Clarendon Press.

Harman, G. (1999b). General foundations versus rational insight. http://www.cogsci.princeton.edu/~ghh/Bonjour.html

Hastie, R. (1993). Inside the juror. Cambridge UK: Cambridge University Press.

Hastie, R., & Dawes, R.N. (2001). Rational choice in an uncertain world : The psychology of judgment and decision making. Sage Publications.

Hintikka, J. (1999). Inquiry as inquiry: A logic of scientific discovery. Dordrecht-Holland: Kluwer Academic Publishers.

Hunter, B. (1992). Criteria and knowledge. In J. Dancy & E. Sosa (Eds.) A companion to epistemology. Oxford UK: Blackwell Publishers.

Ilgen, D.R., Major, D.A., Hollenbeck, J.R., & Sego, D.J. (1995). Raising an individual decision-making model to the team level: A new research model and paradigm. In R.A. Guzzo, E. Salas, & Associates (Eds.), Team effectiveness and decision making in organizations (pp. 113–148). San Francisco, CA: Jossey-Bass

Jackson, S. (1989). What can argumentative practice tell us about argumentation norms? In R. Maier (Ed.), Norms in argumentation. Dordecht-Holland: Foris Publications.

Johnson, R.H. (1996). The rise of informal logic: Essays on argumentation, critical thinking, reasoning and politics. Newport News, VA: Vale Press.

Johnson, R.H. (2000). Manifest rationality: A pragmatic theory of argument. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Johnson, R.H., & Blair, J.A. (1994). Logical self-defense. NY: McGraw-Hill.

Johnson-Laird, P.N. (1983). Mental models. Cambridge, MA: Harvard University Press.

Johnson-Laird, P.N. (1988). The computer and the mind: An introduction to cognitive science. Cambridge MA: Harvard University Press.

Johnson-Laird, P.N. (1993). How the mind thinks. In G. Harman (Ed.), Conceptions of the mind. Mahwah NJ: Lawrence Erlbaum Associates.

Johnson-Laird, P.N., & Byrne, R.M. (1991). Deduction. Mahwah, NJ: Lawrence Erlbaum Associates.

Johnson-Laird, P.N., & Byrne, R.M. (1993). Multiple book review of deduction. Behavioral and Brain Sciences, 16(2), 323-380.

Johnson-Laird, P.N., Legrenzi, P., Girotto, V., Legrenzi, M.S., & Caverni, J.-P. (1999). Naive probability: A mental model theory of extensional reasoning. Psychological Review, 106, 62-88.

Jones, D. (1996). Critical thinking in an online world. http://7-12educators.about.com/gi/dynamic/offsite.htm?site=http%3A%2F%2Fwww.library.ucsb.edu%2Funtangle%2Fjones.html

Kahneman, D. (1976). Attention and effort. Englewood Cliffs, NJ: Prentice-Hall.

Kahneman, D., Slovic, P., & Tversky, A. (1982). Judgment under uncertainty: Heuristics and biases. NY: Cambridge University Press.

Kim, J. (1994). What is "naturalized epistemology"? In H. Kornblith (Ed.), Naturalizing epistemology. Cambridge MA: MIT Press.

Klein, G. (1993). A Recognition-Primed Decision (RPD) model of rapid decision making. In G.A. Klein, J. R. Orasanu, R. Calderwood, & C.E. Zsambok (Eds.), Decision making in action: Models and methods (pp. 138-147). Norwood, NJ: Ablex.

Klein, G. (1998). Sources of Power: How people make decisions. Cambridge, MA: The MIT Press.

Klein, G.A., Orasanu, J.R., Calderwood, R., & Zsambok, C.E. (Eds.) (1993). Decision making in action: Models and methods. Norwood, NJ: Ablex.

Klein, P. (2000). A proposed definition of propositional knowledge. In E. Sosa & J. Kim (Eds.) Epistemology: An anthology. Oxford UK: Blackwell.

Klimoski, R., & Jones, R.G. (1995). Staffing for effective group decision making: Key issues in matching people and teams. In R.A. Guzzo, E. Salas, & Associates (Eds.), Team effectiveness and decision making in organizations (pp. 291-332). San Francisco, CA: Jossey-Bass Publishers.

Kornblith, H. (1994). Naturalizing epistemology. Cambridge MA: MIT Press.

Kuhn, T. (1996). The structure of scientific revolutions. Chicago IL: University of Chicago Press.

Kuhn, D., Amsel, E., & O'Loughlin, M. (1988). The development of scientific thinking skills. San Diego: Academic Press, Inc.

Laudan, L. (1998). Demystifying underdetermination. In M. Curd & J.A. Cover (Eds.), Philosophy of science: The central issues (pp. 320-353). NY: W.W. Norton & Company.

Leddo, J., Cohen, M.S., O'Connor, M.F., Bresnick, T.A., & Marvin, F.F. (1990). Integrated knowledge elicitation and representation framework (Technical Report 90-3). Reston, VA: Decision Science Consortium, Inc.

Lehrer, K. (2000). Theory of knowledge. Boulder, CO: Westview Press.

Lehrer, K., & Paxson, T.D., Jr. (2000). Knowledge: Undefeated justified true belief. In S. Bernecker & F. Dretske (Eds.), Knowledge: Readings in contemporary epistemology. Oxford: Oxford University Press.

Levi, I. (1986). Hard choices: Decision making under unresolved conflict. NY: Cambridge University Press.

Levy, D.A. (1997). Tools of critical thinking: Metathoughts for psychology. Needham Heights MA: Allyn & Bacon.

Lipman, M. (1987). Some thoughts on the foundations of reflective education. In J.B. Baron & R.J. Sternberg (Eds.), Teaching critical thinking skills. NY: W.H. Freeman.

Lipman, M. (1991). Thinking in education. Cambridge, UK: Cambridge University Press.

Lipshitz, R., & Cohen, M.S. (2001). Warrants for prescription: Analytically-based and empirically-based approaches to prescriptive decision making. Manuscript submitted for publication.

Lycan, W.G. (1988). Judgment and justification. Cambridge UK: Cambridge University Press.

Lycan, W.G. (1996). Plantinga and coherentisms. In J.L. Kvanvig (Ed.), Warrant in current epistemology: Essays in honor of Plantinga's theory of knowledge. Lanham MD: Rowman & Littlefield Publishers.

Marr, D. (1982). Vision. NY: W.H. Freeman.

McAllister, J.W. (1996). Beauty and revolution in science. Ithaca, NY: Cornell University Press.

McClelland, J., & Rumelhart, D.E. (1986). A distributed model of human learning and memory. In D.E. Rumelhart, J.L. McClelland, & the PDP Research Group (Eds.), Parallel distributed processing: Explorations in the micro-structure of cognition: Volume 2: Psychological and biological models (pp. 170-215). Cambridge, MA: The MIT Press

McPeck, J.E. (1994). Critical thinking and the "Trivial Pursuit" theory of knowledge. In K.S. Walters (Ed.), Re-thinking reason. Albany, NY: State University of New York Press.

Means, B., Crandall, B., Salas, E., & Jacobs, T.O. (1993). Training decision makers for the real world. In G.A. Klein, J.R. Orasanu, R. Calderwood, & C.E. Zsambok (Eds.), Decision making in action: Models and methods (pp. 306-326). Norwood, NJ: Ablex Publishing Corporation.

Missimer, C. (1994). Why two heads are better than one: Philosophical and pedagogical implications of a social view of critical thinking. In K.S. Walters (Ed.), Re-thinking reason: New perspectives in critical thinking. Albany: State University of New York Press.

Mitroff, I. (1998). Smart thinking for crazy times: The art of solving the right problems. San Francisco: Berret-Koehler Publishers.

Moore, B.N., & Parker, R. (1998). Critical thinking (Fifth Edition). Mountain View CA: Mayfield Publishing Company.

Newell, A. (1981). The knowledge level. Pittsburg, PA: Carnegie Mellon University.

Newell, A. (1990). Unified theories of cognition. Cambridge, MA: Harvard University Press.

Newell, A., & Rosenbloom, P. (1981). Mechanisms of skill acquisition and the law of practice. In J. Anderson (Ed.), Cognitive skills and their acquisition. Hillsdale, NJ: Erlbaum.

Nisbett, R.E. (Ed.) (1993). Rules for reasoning. Mahwah NJ: Lawrence Erlbaum Associates.

Nisbett, R., & Ross, L. (1980). Human inference: Strategies and shortcomings of social judgment. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Nisbett, R.E., & Wilson, T.D. (1977). Telling more than we can know: Verbal reports on mental processes. Psychological Review, 84, 231-259.

Nozick, R. (1981). Philosophical explanations. Cambridge, MA: Harvard University Press.

Oaksford, M., & Chater, N. (1998). Rational models of cognition. Oxford, UK: Oxford University Press.

Orasanu, J. &. Salas, E. (1993). Team decision making in complex environments. In G.A. Klein, J.R. Orasanu, R. Calderwood, & C.E. Zsambok (Eds.), Decision making in action: Models and methods (pp. 327-345). Norwood, NJ: Ablex.,

Oscanyan, F. (1984). Critical thinking: Response to Moore. Teaching Philosophy 7(3), 241-246.

Paris, C.R., Salas, E., & Cannon-Bowers, J.A. (2000). Teamwork in multi-person systems: A review and analysis. Ergonomics, 43, 1052-1075.

Patel, V.L., Arocha, J.F. & Kaufman, D.R. (1999). Expertise and tacit knowledge in medicine. In R.H. Sternberg & J.A. Horvath (Eds.), Tacit knowledge in professional practice. Mahwah NJ: Lawrence Erlbaum Associated, Inc.

Paul, R.W. (1993). Critical thinking: What every person needs to survive in a rapidly changing world. Santa Rosa, CA: Foundation for Critical Thinking.

Payne, J., Bettman, J.R., & Johnson, E.J. (1993). The adaptive decision maker. NY: Cambridge University Press.

Pearl, J. (1989). Probabilistic reasoning in intelligent systems: Networks of plausible inference. San Mateo, CA: Morgan Kaufmann Publishers.

Pearl, J. (2000). Causality: Models, reasoning, and inference. Cambridge UK: Cambridge University Press.

Perelman, C., & Olbrechts-Tyteca, L. (1969). The new rhetoric: A treatise on argumentation. Notre Dame, IN: University of Notre Dame Press.

Plantinga, A. (1993a). Warrant and proper function. NY: Oxford University Press.

Plantinga, A. (1993b). Warrant: The current debate. NY: Oxford University Press.

Pollock, J.L. (1995). Cognitive carpentry: A blueprint for how to build a person. Cambridge MA: MIT Press.

Pollock, J.L. & Cruz, J. (1999). Contemporary theories of knowledge. Lanham, MD: Rowman & Littlefield.

Putnam, H. (1994). Reflexive reflections. In H. Putnam (Ed.), Words and life. Cambridge, MA: Harvard University Press.

Quine, W.V.O. (1953). From a logical point of view. Cambridge, MA: Harvard University Press.

Quine, W.V.O. (1960). Word and object. Cambridge MA: The MIT Press.

Quine, W.V.O. (1970). Philosophy of logic. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Quine, W.V.O. (1990). Three indeterminacies. In R.B. Barrett & R.F. Gibson (Eds.), Perspectives on Quine. Oxford UK: Blackwell.

Quine, W.V.O. (1993). Comment on Katz. In R.B. Barrett & R.F. Gibson (Eds.), Perspective on Quine. Oxford, UK: Blackwell.

Quine, W.V.O. (1994). Epistemology naturalized. In H. Kornblith (Ed.), Naturalizing epistemology. Cambridge, MA: MIT Press.

Quine, W.V. & Ullian, J.S. (1970). The web of belief. NY: Random Nord.

Reiter, R. (1980). A logic for default reasoning. Artificial Intelligence, 13, 81-132.

Rescher, N. (1977). Dialectics: A controversy-oriented approach to the theory of knowledge. Albany: State University of New York Press.

Rieke, R.D. & Sillars, M.O. (1997). Argumentation and critical decision making. NY: Addison-Wesley Educational Publishers.

Rips, L.J. (1994). The psychology of proof: Deductive reasoning in human thinking. Cambridge, MA: MIT Press.

Rubinstein, A. (1998). Modeling bounded rationality. Cambridge, MA: MIT Press.

Schum, D.A. (1994). Evidential foundations of probabilities reasoning. NY: John Wiley & Sons.

Scriven, M. (1976). Reasoning. NY: McGraw-Hill.

Sellars, W. (2000). Does empirical knowledge have a foundation? In E. Sosa & J. Kim (Eds.), Epistemology: An anthology. Oxford UK: Blackwell.

Senge, P. (1990). The fifth discipline. Doubleday. http://www.mapnp.org/library/systems/systems.htm.

Shastri, L. (1999a). Advances in Shruti--A neurally motivated model of relational knowledge representation and rapid inference using temporal synchrony. Applied Intelligence, 11, 79-108.

Shastri, L. (1999b). Types and quantifiers in Shruti--A connectionist model of rapid reasoning and relational processing. In S. Wermter & R. Sun (Eds.), Hybrid neural symbolic integration. Berlin: Springer-Verlag.

Shastri, L., & Ajjanagadde, V. (1993). From simple associations to systematic reasoning: A connectionist representation of rules, variables, and dynamic bindings using temporal synchrony. Brain and Behavioral Sciences, 16, 417-494.

Shepard, R. (1975). Form, formation, and transformation of internal representations. In R. Solso (Ed.), Information processing and cognition: The Loyola Symposium. Hillsdale, NJ: Erlbaum.

Siegel, H. (1997). Rationality redeemed: Further dialogues on an educational ideal. NY: Routledge.

Simon, H.A. (1997). Models of bounded rationality: Empirically grounded economic reason. Cambridge, MA: MIT Press.

Sober, E. (1975). Simplicity. Oxford, UK: Oxford University Press.

Sober, E. (1994). Nature of selection: Evolutionary theory in philosophical focus. Chicago IL: University of Chicago Press.

Soccio, D.J., & Barry, V.E. (1992). Practical logic. Fort Worth, TX: Harcourt Brace Jovanovich..

Sosa, E. (1991). Knowledge in perspective: Selected essays in epistemology. NY: Cambridge University Press.

Sternberg, R.J., & Horvath, J.A. (1999). Tacit knowledge in professional practice. Mahwah NJ: Lawrence Erlbaum Associated, Inc.

Steup, M. (2001). Knowledge, truth, and duty: Essays on epistemic justification, responsibility, and virtue. Oxford UK: Oxford University Press.

Swinburne, R. (1997). Simplicity as evidence of truth. Milwaukee: Marquette University Press.

Thagard, P. (2000). Coherence in thought and action. Cambridge MA: MIT Press.

Thomas, S.N. (1997). Practical reasoning in natural language (Fourth edition). Upper Saddle River NJ: Prentice-Hall.

Tjosvold, D. (1995). Cooperation theory, constructive controversy, and effectiveness: Learning from crisis. In R.A. Guzzo, E. Salas, & Associates (Eds.), Team effectiveness and decision making in organizations (pp. 79–112). San Francisco, CA: Jossey-Bass Publishers.

Toulmin, S.E. (1958). The uses of argument. Cambridge, UK: Cambridge University Press.

Troyer, J. (Ed.) (1998). In defense of radical empiricism: Essays and lectures by Roderick Firth (pp. 205-215). Lanham MA: Rowman & Littlefield.

Unger, P. (2000). An argument for skepticism. In E. Sosa & J. Kim (Eds.), Epistemology: An anthology. Malden MA: Blackwell Publishers.

van Cleve, J. (2000). Foundationalism, epistemic principles, and the Cartesian circle. In E. Sosa & J. Kim (Eds.), Epistemology: An anthology. Malden MA: Blackwell Publishers.

van den Hoven, P. (1987). The external justification of a dialectical consensus. In F.H. van Eemeren, R. Grootendorst et al. (Eds.), Argumentation: Perspectives and approaches. Dordecht-Holland: Foris Publications.

van Eemeren, F.H. & Grootendorst, R. (1992). Argumentation, communication, and fallacies: A pragma-dialectical perspective. Mahwah, NJ: Lawrence Erlbaum Associates.

van Eemeren, F.H. & Grootendorst, R. (1994). Studies in pragma-dialectics. Amsterdam: Vale Press.

van Eemeren, F.H., Grootendorst, R., Henkemans, F. S. (Writers in chief.) (1996). Fundamentals of argumentation theory: A handbook of historical backgrounds and contemporary developments. Mahwah, NJ: Lawrence Erlbaum Associates.

van Eemeren, F.H. Grootendorst, R., Jackson, S., & Jacobs, S. (1993). Reconstructing argumentative discourse. Tuscaloosa AL: University of Alabama Press.

Voss, J.F., Perkins, D.N., & Segal, J.W. (1991). Informal reasoning and education. Mahwah NJ: Lawrence Erlbaum Associates.

Wade, C., & Tavris, C. (1993). Critical & creative thinking: The case of love and war. NY: Harper Collins College Publishers.

Walters, K.S. (1994). Re-thinking reason: New perspectives in critical thinking. Albany: State University of New York Press.

Walton, D.N. (1989). Informal logic: A handbook for critical argumentation. Cambridge, UK: Cambridge University Press.

Walton, D.N. (1995). A pragmatic theory of fallacy. Tuscaloosa AL: University of Alabama Press.

Walton, D.N. (1996a). Argumentation schemes for presumptive reasoning. Mahwah, NJ: Lawrence Erlbaum Associated, Inc.

Walton, D.N. (1996b). Argument structure: A pragmatic theory. Toronto: University of Toronto Press.

Walton, D.N. (1998). The new dialectic: Conversational contexts of argument. Toronto: University of Toronto Press.

Walton, D.N. (1999). Dialectical relevance in persuasion dialogue. Informal Logic, 19, 119-143.

Walton, D.N. & Krabbe, E.C.W. (1995). Commitment in dialogue: Basic concepts of interpersonal reasoning. Albany: State University of New York Press.

Weinberg, S. (1992). Dreams of a final theory. NY: Pantheon Books.

Wooditsch, G.A. (1991). The thoughtful teacher's guide to thinking skills. Mahwah, NJ: Lawrence Erlbaum Associates.

Woods, J. (2000). How philosophical is informal logic. Informal Logic, 20(3), 139-167.

Zsambok, C.E. (1997). Naturalistic decision making: Where are we now? In C.E. Zsambok & G. Klein (Eds.), Naturalistic decision making. Mahwah, NJ: Erlbaum.