A THREE-PART THEORY OF CRITICAL THINKING: DIALOGUE, MENTAL MODELS, AND RELIABILITY[1]

Marvin S. Cohen, Ph.D.
Cognitive Technologies, Inc.
Arlington, VA
www.cog-tech.com
mcohen@cog-tech.com

August 2000

[1] This research was funded by Contract No. DASW01-00-C-3010 with the Army Research Institute, Fort Leavenworth Field Unit. Thanks to Dr. Sharon Riedel for her help throughout this project.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
AUG 2000		**00-00-2000 to 00-00-2000**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
A Three-part Theory of Critical Thinking: Dialogue, Mental Models, and Reliability	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Cognitive Technologies Inc,4200 Lorcom Lane,Arlington,VA,22207**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
Proceedings of Army Research Institute Critical Thinking Conference, December 2001. U.S. Government or Federal Rights License

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**26**	

Should the Army be interested in critical thinking?

Is critical thinking important? And if so, why? A small set of themes appear over and over in the prefaces and introductions of the dozens of critical thinking textbooks that are in print. Claims fall into three groups: A. *Problem difficulty*, including (i) increasing complexity of problems, (ii) changing nature of problems, and (iii) information overload. B. *Decentralized social and organizational structure*, including (i) increasing responsibility and need for initiative, (ii) increasing participation in teams with diverse membership, and (iii) increasing need for independent thinking. C. *High stakes*, including (i) important public policy issues and (ii) personal decisions in an increasingly competitive career environment.

Do conditions for the use of critical thinking apply in the Army? The answer certainly appears to be yes. There is a growing interest in critical thinking among Army instructors and researchers, which seems warranted by (A) the complexity and changing character of military planning and operations; (B) decentralization of the organizational structure (e.g., the demands of leadership, coordination, and initiative within every echelon); and (C) high stakes personally, organizationally, and for the nation as a whole. In addition, the direction of change in the Army promises to make critical thinking even more important. 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 military organizations, and the growing interest in tactics that rely on initiative by local commanders.

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 the rhetorical compatibility of claims for critical thinking and Army needs – and to evaluate the prospects of a match at a deeper and more detailed level. To dramatize the need for clarification and coherence, let us play devil's advocate. The current state of critical thinking research and instruction leaves unanswered some important questions about the application of critical thinking to the Army tactical battlefield domain:

1. 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"?
2. 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?
3. 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 within diverse, multi-national operations because it is "Western, masculine, individualistic, adversarial, and coldly rational" (Atkinson, 1998: p.121).
4. 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?

This chapter can only scratch the surface in trying to respond to these challenges. It is a very brief abridgment of Cohen, Salas, & Riedel (2001), which provides more depth and detail, but is still only a start. The research had two main goals: First, to draw a map that links disparate regions of the critical thinking field, and second, to use the map to navigate toward a more insightful theory of critical thinking, which will support the development more effective methods for improving it in Army battlefield command teams.

Three components of critical thinking

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 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) cognitive theories of reasoning according to which alternative possible situations are represented by mental models; (2) normative models of critical discussion in which a proponent must defend a claim against an opponent or critic; and (3) models of cognitive mechanisms and of the environment which enable us to assess the reliability of the processes by means of which we form beliefs and make choices.

Critical thinking, like an onion, has a multi-layered structure (Figure 1). 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 selective consideration of *alternative possible states of affairs*. Metrics of performance at this level involve logical, probabilistic, and explanatory coherence of mental models.
2. At the intermediate level, these models are embedded within a layer of *critical questioning* which motivates the generation and evaluation of possibilities. Such dialogues may take place within a single individual, or they may be conducted among different individuals. Critical questioning is evaluated by reference to norms for conducting the appropriate kinds of critical dialogue. Dialogue types are differentiated by the depth of probing to which a proponent must respond and the scope of the permitted responses.
3. At the outermost layer, critical thinking is a judgment about the *reliability* of a cognitive faculty, hence, the degree of *trust* that should be placed in its outputs. The critical dialogue is only one of various available cognitive or social processes that might be utilized to generate beliefs and decisions. Different processes, such as pattern recognition, may be more reliable under some conditions.

In sum, critical thinking skill is exemplified by *asking and answering critical questions about alternative possible states of affairs, to the extent that such questioning is likely to increase the reliability of the overall activity in achieving its purpose*.

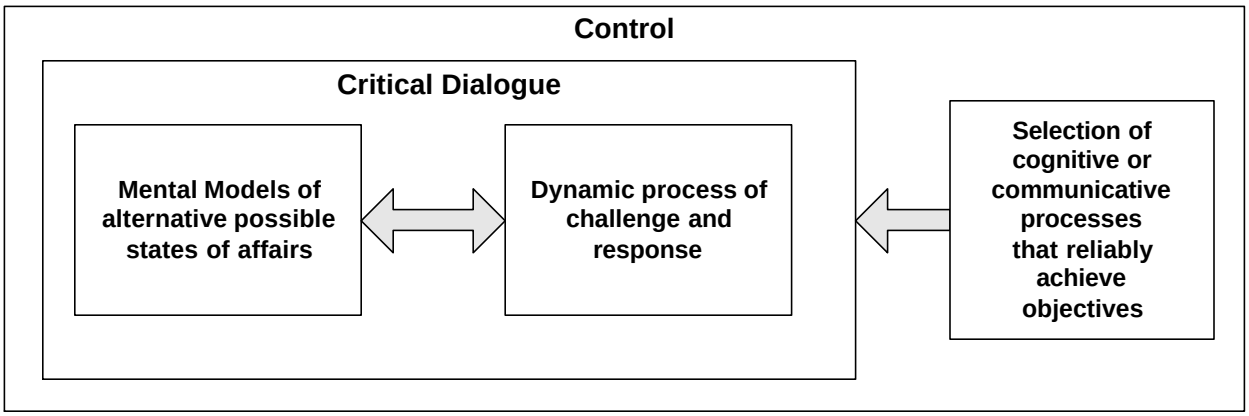

Figure 1. A model of critical thinking with three embedded layers: mental models, critical dialogue, and control based on reliability.

In the remainder of this chapter, we will very briefly discuss some of the background and rationale for this theory, and return at the end to the question of usefulness in Army battlefield decisions.

Avoiding the pitfalls of intellectualism

Modern philosophy began (e.g., Descartes, Locke, Hume) with the notion that we have a duty to carefully *decide* whether to accept or reject our beliefs, and a duty to base those decisions upon good *evidence*. At the beginning, philosophers thought evidence was good only if it rendered a conclusion absolutely certain. Now, philosophers acknowledge uncertainty, and good evidence only needs to provide sufficient justification (whatever that is). But even so, a consistent underlying point of view has persisted and has had an enormous influence on the critical thinking movement. Even if our beliefs are *true*, unless we accepted them on the basis of what we explicitly take to be good evidence, we are correct only by a lucky accident (P. Klein, 2000).

Example

Suppose MAJ Jones correctly believes that there is an enemy T-62 tank in the vicinity. Suppose that she believes it because she saw the tank and is in fact highly accurate in recognizing types of tanks. MAJ Jones, however, does not believe that she is sufficiently skilled to identify a T-62. Does MAJ Jones *know* that the tank is a T-62 under these circumstances? The intellectualist viewpoint would say no. Even though she accepts the belief based on good evidence, she does not *take* the evidence to be good. Thus, she was right about the tank by accident

From this point of view, the purpose of critical thinking is to ensure that we have explicit reflective knowledge of all our first-level beliefs, our reasons for accepting them, and the criteria that determine whether the reasons are sufficient. Sosa (1991, p. 195) dubbed this view the "intellectualist model of justification." But is this the best view of what critical thinking is all about?

Siegel (1997) falls well within the intellectualist 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 seems to be nothing more than simple common sense. Not surprisingly, the notion of argument (i.e., reasons for conclusions) is central in textbooks and theoretical discussions of critical thinking. Unfortunately, if applied universally and consistently as Siegel (1997: p. 16) says it should be, the demand for argument raises the danger of an infinite regress (Dancy & Sosa, 1992: p. 209-212). If reasons are required for every belief,

then reasons must be provided to justify the reasons, to justify the reasons of those reasons, and so on. Critical thinking may never come to an end. There are only three ways to respond to this problem within the philosophical tradition: The list of reasons is infinite, it circles back on itself, or it stops. If it stops, the point at which it stops may be arbitrary or non-arbitrary.

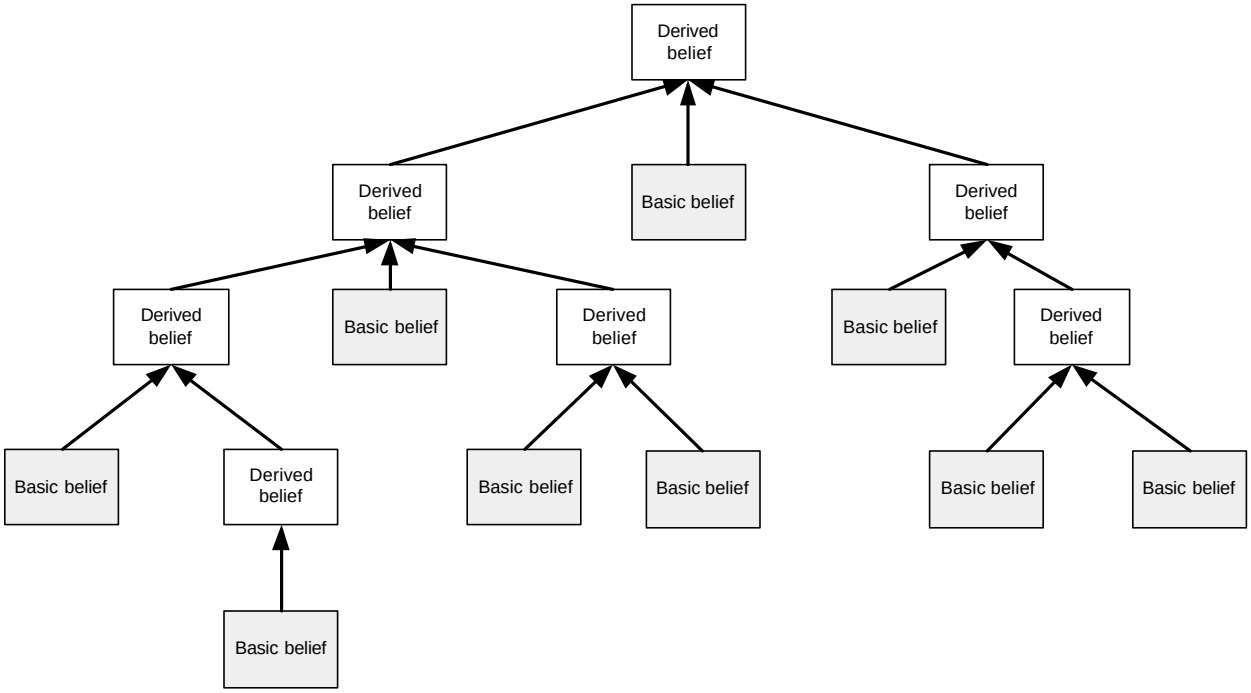

Figure 2. Foundationalist paradigm for acceptability of beliefs: a pyramid. Arrows represent arguments in which conclusions are inferred from reasons. Every chain of argument must be traceable back to basic beliefs, which are anchored in experience or logical intuition (shaded boxes).

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., Ungar, 1974; Foley, 1990). A variant of skepticism is that the list of reasons ends with arbitrary assumptions that provide the basis for the other beliefs. This is the *relativist* position, that beliefs are not justified absolutely, but only relative to the framework of assumptions that happens to be accepted in a domain or culture, or even by a specific individual in a specific context. But it is hard to make a case for the usefulness of critical thinking if there are no intersubjective standards of reasoning.

A second possibility is that the chain of reasons eventually circles back on itself. For example, continuing down the chain of reasons, we would eventually arrive again at the conclusion. Siegel, like many critical thinking theorists and informal logicians, rejects the idea that a chain of arguments can legitimately circle back on itself. As he sees it (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. Siegel (along with most other theorists in critical thinking and informal logic) is therefore committed to the third and most ambitious possibility, that the list of reasons must come to rest on solid ground, with beliefs that do not themselves require reasons and which can serve as foundations for other beliefs (Figure 2). These beliefs must be distinguished by some intrinsic cognitively accessible feature that lends them a higher level of certainty, such as their origins in perception or logic. That view is called *foundationalism* (Chisholm, 1977; Pollock & Cruz, 1999).

Coherence

Unfortunately, foundationalists have been unable to successfully define a convincing class of basic beliefs for which arguments are unnecessary. Virtually every belief depends in some way on other beliefs for its justification. For that reason and others, many philosophers urge consideration of a more sophisticated variant of the "circular reasoning" option called *coherentism* (Thagard, 2000; BonJour, 1985; Lehrer, 2000; Harman, 1986; Quine & Ullian, 1970; Everitt & Fisher, 1995). Coherentists accept that a chain of arguments for a conclusion will, if pursued long enough, arrive back at the conclusion itself, just as a chain of dictionary definitions will eventually arrive back at the original word. An explanatory hypothesis draws support from the observations that it explains, but also, the veracity of the observations is supported by the existence of a good explanation. In short, the premises in an argument are not "basic" in any deep sense that differentiates them from the conclusion. "Arguments" might run in either direction.

Example

Suppose MAJ Smith 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 Smith 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 Smith to question the map or the reports of dummy tanks. MAJ Smith 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 Smith 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 of which is justified with respect to the others by the same set of coherence relationships.

Reasoning may be circular if carried on long enough, but coherentists deny that *justification* is circular because they reject the foundationalist equation of justification with reasoning. Justification is not directly transferred from one belief to another by a linear series of *arguments* (Day, 1989). From the coherentist perspective, it is the *system* of beliefs that is the target of justification, not the individual beliefs within it (Figure 3). A system of beliefs is coherent when its members are tightly interconnected by logical, conceptual, explanatory, or other such relationships. Every belief contributes some support to every other belief and in turn draws support from every other belief, just as each stone in an arch depends on the other stones. 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 because reasons *could* be given, e.g., by citing 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 clearly, arguments for individual beliefs have a much diminished role in settling questions of justification.

We have seen that a central problem of critical thinking is how to know when to *stop* demanding reasons for a belief. Some possible answers are:

- Skepticism: Never – justification cannot be completed.
- Relativism: At assumptions that cannot themselves be justified.
- Foundationalism: At a rock-bottom set of beliefs, based on sense perception or logic, that do not require justification in terms of other beliefs.

- Coherentism: At any already accepted members of a coherent system of beliefs.

None of these positions is altogether satisfactory. On the one hand, as we mentioned, foundationalism fails because all beliefs depend on other beliefs. A second problem with foundationalism is that it fails to explain how to choose between plausible arguments that lead to conflicting conclusions. (This is also a problem with informal logic, as discussed in Cohen, Salas, & Riedel, 2001.) On the other hand, coherentism has the opposite problems: First, some beliefs do in fact receive priority over others, even in they are not known with certainty; examples include observation reports about nearby objects in plain sight. But how can this priority be explained if every belief is justified in the same way, by its place within the same system of beliefs? Second, coherentism provides an account of how conflicting arguments are resolved, via the evaluation of alternative systems of belief. But even for moderately sized belief systems, the combinatorics of inference far exceed human cognitive capabilities (Cherniak, 1986).

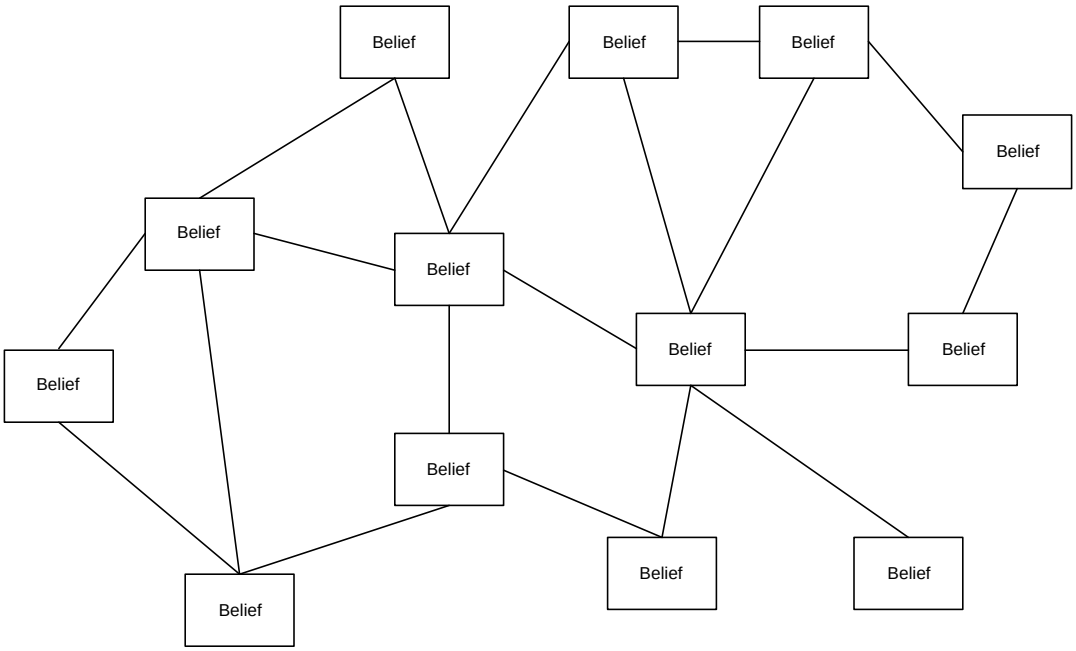

Figure 3. Coherentist paradigm for acceptability of beliefs: a network. The system of beliefs is justified as a whole by the inferential links among its components and its overall simplicity and comprehensiveness. Beliefs are not classified into types with different epistemological status, such as basic or not basic.

There are two kinds of responses to problems with pure coherentism, one of which makes intellectualism worse, while the other makes a dramatic break with it. The intellectualist "solution" (BonJour, 1985; Lehrer, 2000; Harman, 1973) requires a higher degree of reflective self-awareness, stipulating that the system of beliefs be evaluated in terms of coherence with second-tier beliefs about the origins and reliability of all its first-order beliefs. Thus, the priority of perceptually based beliefs is a consequence of the coherence of meta-beliefs about how reliable perception is under relevant conditions. This requirement constitutes an admission that pure coherence is insufficient to support the justification of beliefs. Specific *kinds* of beliefs (i.e., second-order beliefs about reliability) must be part of the mix. However, the demand for continuous reflective awareness exceeds human capabilities. It also threatens another kind of vicious regress, involving beliefs about beliefs, beliefs about those beliefs, and so on, unless it reverts to a form of foundationalism in which beliefs about reliability are the unquestioned foundations in need of no further arguments (Sosa, 1991, pp. 205-207).

Reliability

The more appealing solution is to accept that some of the factors justifying belief acceptance may not be cognitively accessible. First, perceptual systems may reliably anchor a system of beliefs in reality even if the subject has no explicit reflective awareness of their reliability. Second, coherence may be established by relatively automatic processes of spreading activation across a network of beliefs, rather than as a result of deliberate reasoning (Thagard, 2000; Cohen, Thompson, Shastri, Salas, Freeman, Adelman, 2000b). The mutual influence of beliefs (hence, the scope of coherence) will depend on the distance that activation must travel in the network. The role of deliberate critical thought, on the other hand, is more limited: It will selectively activate and evaluate modular *subsets* of beliefs, i.e., mental models (Kornblith, 1989; Cohen, Thompson, Shastri, Salas, Freeman, Adelman, 2000a). These responses imply that a belief, whether perceptual or inferential, may constitute genuine knowledge even though the cognizer is unable to articulate reasons for holding it. This idea is called *externalism.* (The idea that our evidence for beliefs must be conscious or readily made conscious is called *internalism*, and is shared by both foundationalism and coherentism.)

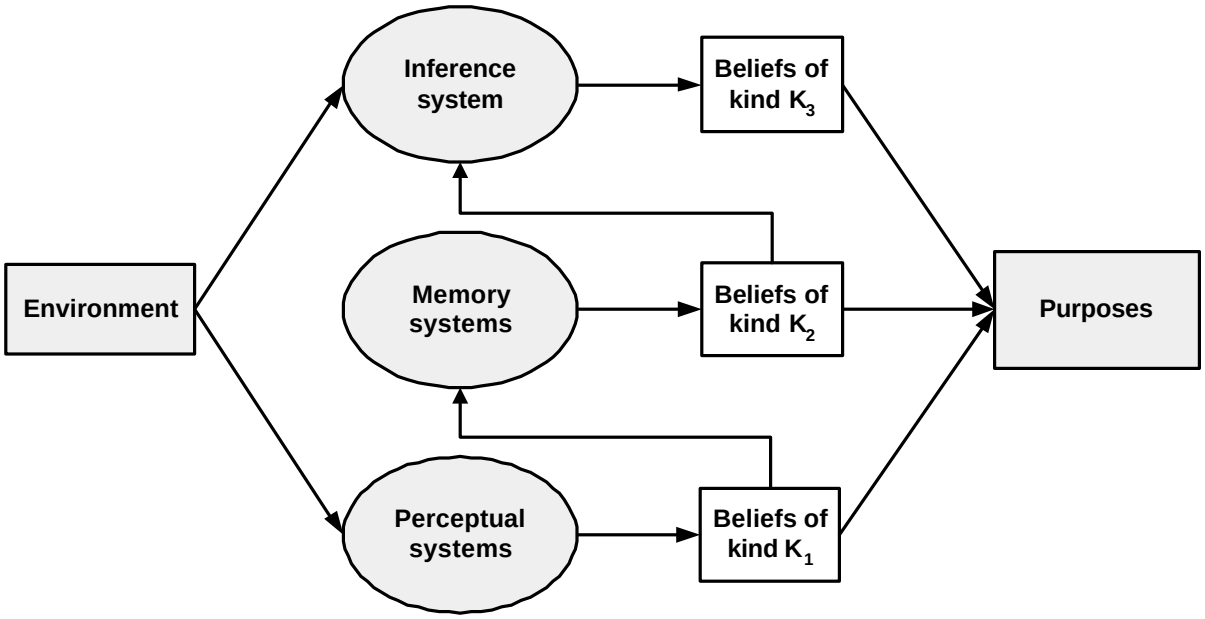

Figure 4. Reliabilist paradigm for acceptability of beliefs: a series of input/output processes. Beliefs are justified to the extent that they are produced or sustained by processes that reliably achieve the goal of accepting true beliefs and avoiding false beliefs under relevant environmental conditions. 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.

Externalism has attracted considerable recent interest from philosophers (e.g., Goldman, 1992, 1986; Dretske, 1983; Nozick, 1981; Sosa, 1991; Plantinga, 1993). According to one version of externalism, called reliabilism (Nozick, 1981), a belief is justified if it is generated or sustained by a method that "tracks" the truth, i.e., tends to produce the belief that *p* if and only if *p* is true. According to another variant of 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. 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 (Figure 4).

Externalism does not insist that a person have cognitive access to reasons for a belief, that a person have second-order beliefs about the reliability of first-order beliefs, or even that beliefs are always

under voluntary control. A person is deemed expert or non-expert based on performance and results: the actual accuracy of her 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.

Objections to reliabilism have stressed several points: First, there is the *coherence* problem. 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?

Second, there is the *generality* problem. The reliability of a cognitive faculty might be thought of as 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, then 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 third and final problem concerns *fairness* in evaluation. Recall MAJ Jones, who has a highly reliable faculty for quickly recognizing different types of tanks as a result of long training and experience. But MAJ Jones does not realize how reliable her judgment is and indeed believes it to be unreliable. MAJ Jones would seem to be unjustified in accepting her own beliefs about tanks, even thought they are reliable (Bonjour, 1985). Given her beliefs about her own unreliability, she would be right to double check the tank identifications before accepting them. Conversely, recall MAJ Smith. Her faculty for recognizing the presence of a tank is generally reliable, but is unreliable under special circumstances (such as when there are dummy tanks in the area). But if MAJ Smith had no way of knowing that dummy tanks were in the area, or indeed had reason to believe there were none, wouldn't her tank identifications be justified even though they were unreliable? Both of these points have been taken to suggest that internalist intuitions based on fairness, both in holding people responsible for errors and giving them credit for successes, are not accounted for by externalism.

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 MAJ 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 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. Dual-perspective reliabilism takes seriously the coherentist conclusion: 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 Smith's perceptual recognition of a tank and is aware of the presence of dummy tanks in the area, J will not regard MAJ Smith's judgment as reliable evidence for the presence of a tank. But if J trusts MAJ Smith 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 Smith's tank report is reliable.

The fairness problem is in part 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 cognitive 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 assessor J and of 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. From MAJ Smith's own point of view, her recognition of a tank is reliable because she believes that dummy tanks are not likely. And since MAJ Jones thought her tank identifications were unreliable, she could not be blamed for seeking further verification before accepting her perceptual judgments. Both MAJ 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 had 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 implementers, 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 prepare for potential problems during the execution of the plan. But since the implementers inevitably overlooked some possibilities, they overestimated overall chance of success. When participants performed as plan developers, 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 enumerating failure scenarios. As a result, planners 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 itself 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.

Reliability and coherence in 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.

The three-part model of critical thinking (Figure 1) integrates insights from coherentist and reliabilist theories of justification. The version of reliabilism depicted in Figure 4 has a foundationalist flavor because reasoning builds on a distinct, privileged class of beliefs generated by perception. By contrast, Figure 5 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 5 rejects the foundationalist assumption that there is a privileged class of beliefs that is immune to reasoning. At the same time, 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 report 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 perceptual beliefs.

The three-part model of critical thinking forms the top tier of Figure 5, 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.

We will now discuss in a bit more detail how the components of this model work together in critical thinking.

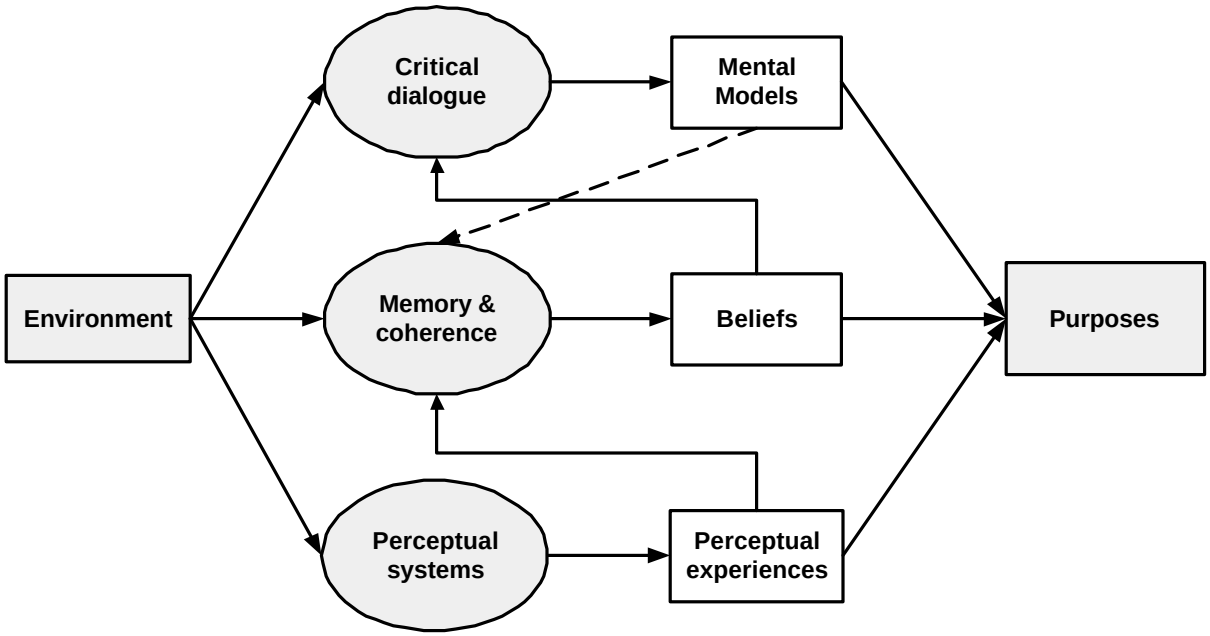

Figure 5. A reliabilist framework that integrates a coherence theory of reasoning with the critical thinking model in Figure 1.

Mental Models

Johnson-Laird (1983; Johnson-Laird & Byrne, 1991) cites evidence that humans reason not (or not only) in terms of syntactic formal patterns but in terms of *meaning*. Comprehending an assertion includes understanding what possible states of affairs are compatible with the assertion and which are excluded (Johnson-Laird & Byrne, 1991). *Inference* is in large part a process for comprehending multiple assertions, that is, for determining what states of affairs are consistent with several different assertions. The conclusion of an inference must be true in every surviving possibility.

The representation of a possible state of affairs is called a *mental model*. Typically, a simple statement (e.g., *the enemy will attack through the northern pass*) is compatible with many different states of affairs (when and how they will attack), but people often use a single, representative mental model to conserve processing capacity (Johnson-Laird & Byrne, 1991, p. 170). They change the representation or expand it to include other possibilities only when forced to do so. When a sentence contains logical connectives, such as *and*, *or*, *all*, and *some*, people use knowledge of the meaning of such connectives to construct some or all of the appropriate mental models. Suppose MAJ House and MAJ Kerr are analyzing enemy intent, and MAJ House says:

MAJ House: *The enemy will attack either through the northern pass or the southern pass.*

Neither MAJ House nor MAJ Kerr believes the enemy will attack through both passes. The *or* statement thus suggests only two possible mental models: one in which the enemy attacks in the south (and not the north), and the other in which they attack in the north (and not the south):

1	*The enemy will attack through the south*	*NOT: The enemy will attack through the north.*
2	*NOT: The enemy will attack through the south*	*The enemy will attack through the north.*

To conserve processing capacity, according to Johnson-Laird, and Byrne, people usually do not represent negations unless they are stated explicitly. So, the two cells with NOT would be left blank.

Now suppose MAJ Kerr says, "*I don't believe it will be the northern pass."* MAJ Kerr has committed herself to mental model #1:

1	*The enemy will attack through the south.*	

MAJ House, however, is still uncertain, and wants to know the reasons for this conclusion. MAJ Kerr replies, " *Because they don't have any artillery there."* Background knowledge of enemy doctrine suggests that the enemy will not attack without using artillery first to soften up the opposing force. Thus, only attack in the south is consistent with the absence of artillery in the north. Hence, MAJ Kerr has concluded that the enemy will not attack in the north. The following mental models capture MAJ Kerr's reasoning, and also show how she intends to persuade MAJ House to accept mental model #1:

	The issue		**MAJ Kerr's reason**	
1	*The enemy will attack through the south*		*NOT: The enemy has artillery in the north.*	This mental model is consistent with background knowledge.
2		*The enemy will attack through the north.*	*NOT: The enemy has artillery in the north.*	This mental model is ruled out by background knowledge of likely enemy tactics.

In this example, background knowledge ruled out mental model #2. However, Johnson-Laird and Byrne (1991) show how such conclusions can be reached by more explicit deductive reasoning with mental models. Such reasoning is typically more effortful. For example, suppose MAJ Kerr had said:

> MAJ Kerr: *If they were going to attack in the north, they would have artillery nearby. And they don't.*

Considering only the simple components of MAJ Kerr's statements, there are four logical possibilities, corresponding to different combinations of truth and falsity of attack in the north and artillery in the north. The conditional statement (*If they were going to attack in the north, they would have artillery nearby*) excludes just one of these four situations:

1	*NOT: The enemy will attack through the north.*	*NOT: The enemy has artillery in the north.*	
2	*The enemy will attack through the north.*	*NOT: The enemy has artillery in the north.*	This situation is excluded by the If__then__ statement.
3	*NOT: The enemy will attack through the north.*	*The enemy has artillery in the north.*	
4	*The enemy will attack through the north.*	*The enemy has artillery in the north.*	

(We omit the column for attack in the south for simplicity.) The other part of MAJ Kerr's argument is that the enemy does not have artillery in the north. Adding that piece of information excludes mental model #3 and mental model #4. Thus, the only surviving possibility is mental model #1, and the conclusion is that the enemy will not attack in the north.

According to mental model theory, the difficulty of an inference increases with the number of alternative possibilities that must be considered. Thus, the use of background knowledge (i.e., the automatic operation of a coherence-based network of beliefs) to eliminate at least some of the possibilities, as in our example, is much less effortful than explicit inference. Indeed, a major advantage of mental model theory over other approaches is that it can accommodate both automatic and deliberate processes in any mix. Errors in explicit inference 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 explicit and true components of premises and thus to neglect possibilities consisting of false components; or a prior tendency to believe that the conclusion is correct causes the reasoner to cut short the exploration of alternatives. Of course, another possible cause of error is elimination of possibilities due to inaccurate background beliefs. 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.

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).

The pragma-dialectical theory proposed by van Eemeren and Grootendorst (1994) closely interweaves normative and descriptive elements. An ideal of critical rationality in dialogue is developed, while 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, & Snoeck 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 (1996b), "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…."

Dialogue theory provides a deeper analysis of fallacies than the usual description in terms of surface features. For example, one fallacy is typically described as attacking the person, or *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." *Ad hominem* fallacies sometimes involve violation of this rule. 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 is what makes this an error, not surface features ("attacking the person"). Other fallacies (e.g., argument by appeal to pity, or threats of force) that described very differently from *ad hominem* appear to involve violation of the same dialogue principle and thus are the same error when considered at a deeper

level. Conversely, in other contexts, impugning the character of a person may be highly appropriate, e.g., if the person's testimony must be relied on in drawing a conclusion. Understanding errors in terms of dialogue rules provides both a more nuanced and a more accurate assessment of their normative status.

Walton (1998) has studied a variety of different kinds of dialogue, which differ in their purposes and the norms by which they are conducted: e.g., deliberation, inquiry, negotiation, information seeking, and persuasion. 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 reflective 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).

According to van Eemeren and Grootendorst (1992, pp. 34-37), a *critical discussion* is a dialogue type used for the resolution of a difference of opinion. Resolution is not a matter of negotiation (which is a different type of dialogue) or of simply setting the difference aside. It involves *persuading* one of the parties to retract doubt concerning the other party's position because she has been convinced by the other party's reasons, or conversely for one of the parties to relinquish her own position because it has not withstood the other party's challenges. In some dialogue models (e.g., Walton, 1998; van Eemeren and Grootendorst, 1992), there are two participants or roles: a *proponent* and an *opponent*. In other models (e.g., Rescher, 1977), there are three participants or roles: a *proponent*, an *opponent*, and a *judge*. Rules governing the possible actions of each participant are a function of the type of dialogue, the stage of the dialogue, and the previous statements of each participant.

In the *confrontation* stage of a critical discussion, a difference of opinion is acknowledged. For example, the proponent expresses a standpoint with or without reasons, and the opponent indicates disagreement or expresses doubt. The parties may also seek to clarify or flesh out each other's positions. In the *opening* stage, which is likely to be implicit rather than explicit (van Eemeren and Grootendorst, 1992, p. 41), the parties "agree" on the type of discussion they will have and the discussion rules. Specifically, in the type of dialogue called a critical discussion, they agree that one will take the role of proponent and the other will take the role of the opponent. 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. They agree that each assertion must support the goal of the dialogue type they have selected, e.g., to resolve the difference of opinion, and they agree not to shift dialogue types without mutual agreement. They also agree to distinguish which assertions are meant as conclusions and which are meant to be reasons for those conclusions.

The crucial stage of a critical discussion is *argumentation*, in which the proponent and opponent carry out their roles of defending and challenging a thesis, respectively. Normative models of this stage spell out the types of assertions that are permitted to each side as a function of previous assertions. The major difference between the proponent and opponent in a critical dialogue is the global burden of proof. It is up to the proponent to create a positive case for her standpoint. The opponent merely has to create doubt. (In more complex types of dialogue, the two parties may defend contrary theses, and each participant in effect plays opponent to the other.) 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. In the *concluding* stage of a critical discussion, the dispute may be ended because the proponent withdraws her thesis or because the opponent withdraws her doubt.

The critical discussion (or the more general *persuasion* dialogue described by Walton, 1998) 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 idea of 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 the functional resemblance between thought and dialogue is more than a coincidence. A variety of developmental psychologists (starting perhaps with Vygotsky) have proposed that thought 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.

Mental models and dialogues

The argumentation stage of a critical dialogue can be seen as a process of constructing and evaluating mental models. 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 & Krabbe (1995) distinguish two kinds of explicit commitment stores: commitments based on assertions, which the speaker is obligated to defend, and commitments by a listener based merely on concessions, which the listener 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. The listener can challenge any assertion by the speaker as long as that assertion is not in the listener's own assertion-based commitment store. If the listener challenges a commitment based on an assertion by a speaker and the speaker cannot defend it by supplying reasons, the speaker must retract it. When the listener does not immediately challenge an assertion by the speaker, the listener has conceded it, and it goes into the listener's concession-based commitment store. The listener is of course not obligated to defend her concession, but must allow the speaker to use it in argumentation at least for the time being. The listener can retract the concession at any time simply by challenging it, as long as it is still in the speaker's commitment store (otherwise, the challenge would be irrelevant). The speaker can also choose to retract an assertion of her own, but this is more difficult to do because she must also find and retract any other commitments that imply the retracted assertion (i.e., the reasons she may have given for her assertion). If there are inconsistent assertions in the speaker's commitment store, and the listener challenges them, then the speaker must retract at least one of the conflicting commitments along with the reasons that led to it.

Commitment stores are simply sets of mental models. Each mental model in the commitment store of a dialogue participant represents a state of affairs that is regarded as *possible* by that participant at that particular time. For example, in the argument about location of attack, MAJ Kerr is the proponent. She stated a thesis (that the enemy attack will be in the south) which MAJ House did not concede. MAJ House thus takes the role of opponent. In response to MAJ House's challenge, MAJ Kerr gave a reason (lack of artillery in the north) that was intended to persuade MAJ House to exclude the competing possibility.

Now suppose MAJ House concedes that absence of artillery in the north would be a good reason to accept MAJ Kerr's conclusion, if the reason were true. (In other words, she chooses not to challenge

that aspect of the argument for the time being.) But MAJ House expresses doubt about the reason itself: "*What makes you think there is no artillery in the north?*" Since MAJ House does not concede the truth of the reason, the opponent's mental models now include the following different situations:

Opponent

	The issue		Proponent's reason	
1	*The enemy will attack through the south*		*NOT: The enemy has artillery in the north.*	
2		*The enemy will attack through the north.*	*NOT: The enemy has artillery in the north.*	Opponent concedes that this possibility is (temporarily) excluded
3	*The enemy will attack through the south*		*The enemy has artillery in the north.*	Opponent questions reason, hence, still believes these are possible.
4		*The enemy will attack through the north.*	*The enemy has artillery in the north.*	

Notice that the opponent, MAJ House, is not committed to the negation of the reason (that there *is* artillery in the north), but only to the *possibility* that there is artillery in the north. Nor is the opponent saying that the conclusion is false. But she is saying that the possibility of artillery in the north opens up the *possibility* of attack in the north. Specifically, in mental model #4 the enemy has artillery in the north and attacks through the north. Since the proponent's conclusion is not true in all the possible situations, the opponent is not yet convinced. In order to persuade the opponent, the proponent must find some way to eliminate mental model #4. Suppose that the proponent now presents a reason to believe that the enemy has no artillery in the north:

Proponent

	The issue		Proponent's reason	Proponent's reason for the reason
1	*The enemy will attack through the south.*		*The enemy has no artillery in the north.*	*Our imagery of the northern sector is excellent and shows no artillery in the north.*

This time the opponent, MAJ House, concedes the truth of MAJ Kerr's reason (the imagery is good and showed no artillery in the north). But now she wishes to go back and challenge something that she had temporarily conceded: MAJ Kerr's claim that absence of artillery is in fact a good reason under these circumstances for expecting no attack in the north. In particular, MAJ House points out, "*Don't we have reports that the enemy has developed longer-range artillery?*" This move is the opposite of the previous one. MAJ House accepts the truth of the reason (no artillery in the north) but has introduced a consideration intended to neutralize or cancel it out as evidence for the conclusion (no attack in the north). The absence of artillery in the north, in conjunction with the fact that the enemy has developed longer range artillery is *not* evidence for the conclusion (since the enemy could use artillery located at a longer distance). We call such an objection a *defeater*. The opponent's mental models are now the following:

Opponent

	The issue		Proponent's reason	Proponent's reason for the reason	Opponent's defeater
1	*The enemy will attack through the south.*		*NOT: The enemy has artillery in the north.*	*Our imagery of the northern sector is excellent and shows no artillery in the north.*	*The enemy has developed long range artillery.*
2		*The enemy will attack through the north.*	*NOT: The enemy has artillery in the north.*	*Our imagery of the northern sector is excellent and shows no artillery in the north.*	*The enemy has developed long range artillery.*
			…	…	

The three dots indicate that the opponent has merely conceded the imagery evidence and thus the absence of artillery in the north, but is not committed to defending them. She is at liberty to challenge them again later. We have used Johnson-Laird's (1983) convention for representing implicit mental models as a handy way to represent concessions in a commitment store.

The opponent's challenge to MAJ Kerr's argument introduces the very important topic of *defeasibility*, and shows how dialogue theory and mental models in conjunction help clarify some key aspects of reasoning about uncertainty. Informal logicians, psychologists, philosophers, and artificial intelligence researchers generally agree that non-deductive inferential conclusions are subject to defeat by new information, i.e., such inferences are *defeasible*. As we have seen, a defeater (e.g., the development of longer range artillery) may undermine an inference without providing evidence for the opposite conclusion. Notice in addition that the opponent does not even deny that the original evidence (the absence of artillery in the north) supported the proponent's conclusion (no attack in the north). The opponent merely points out that while no artillery generally indicates no attack, there are special circumstances in this situation that must be taken into account. The further bit of information about longer range artillery *neutralizes* the support given by the proponent's evidence for the proponent's conclusion in this context. At this stage of the dialogue, MAJ Kerr's argument is defeated and no conclusion about location of attack can be drawn. Thus, she may have to retract her conclusion about attack in the south.

Defeasibility, however, is an open-ended aspect of reasoning about the real world. Thus, MAJ Kerr may answer MAJ House's challenge by defeating the defeater. For example, MAJ Kerr replies: "*That may well be, but I don't recall any indications that they've deployed the new systems yet.*" If the enemy has not deployed the new artillery, then mere development of the technology is irrelevant. The original argument based on lack of artillery in the north regains its former force. The proponent's commitment store still has only one explicit mental model:

Proponent

	The issue		Proponent's reason	Proponent's reason for the reason	Opponent's defeater	Proponent's defeater of the defeater
1	*The enemy will attack through the south.*		*NOT: The enemy has artillery in the north.*	*Our imagery of the northern sector is excellent and shows no artillery in the north.*	*The enemy has developed long range artillery.*	*The enemy has not deployed the new artillery.*
					…	

The three dots show that MAJ Kerr has conceded the development of longer range artillery but may choose to make alternative possibilities explicit later. In this mental model, it is clear that the three claims in combination – lack of artillery in the north, and possession of longer range artillery that has not been deployed – do provide evidence against attack in the north. Thus, only one possibility survives, mental model #1, in which the enemy attacks in the south. MAJ Kerr's original conclusion has been vindicated – unless of course the opponent comes up with another challenge, to which MAJ Kerr has no response.

Defeasibility is pervasive in everyday reasoning but is not handled well within either formal or informal logical systems. Logicians tend to deal with defeasibility by tinkering with the premises or inference rules of a reasoning system. For example, they might add the falsity of the defeater to the premises in the argument – e.g., artillery location is an indicator of location of attack only if longer range artillery has not been developed and deployed. The problem with this tactic, aside from computational complexity, is that it blocks reasoning with incomplete information. The falsity of all possible defeaters would have to be positively determined whenever artillery was used as an indicator of location of attack. But in many circumstances, this is either not possible or not worth the time. As the conversation between MAJ Kerr and MAJ House continues, more exceptions and exceptions to exceptions may be brought forward. Each new addition of clauses to the premises would ratchet 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 partial solution is to add special default inference rules, so that the conclusion follows in the absence of positive evidence that the defeaters are true (e.g., Reiter, 1980). Again, however, this introduces extreme computational complexity. Another problem is that the list of potential defeaters is indefinitely long, and advance specification of all defeaters in special default rules 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. Even more importantly, the logical approaches provide neither guidance nor 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 initially incomplete set of mental models. Put the other way, it involves the discovery of possible states of affairs that were not previously considered but which are relevant to the conclusion in the current context. Thus, it lends itself to a semantic mental model-based approach that represents the alternative possibilities that are considered in reasoning (Johnson-Laird & Byrne, 1991; Johnson-Laird, Legrenzi, Girotto, Legrenzi, & Caverni, 1999). The reasoning process itself alternates steps of generating new possibilities and using background

knowledge or explicit inference to evaluate their plausibility. Dialogue theory provides norms for the process of challenge and response during which mental models are elaborated and accepted or rejected. Finally, as we shall see, judgments of reliability determine what process should be used and when the process should stop in any particular situation

As the critical dialogue progresses, new features are added to the model, either to challenge or to defend the proponent's conclusion. Thus, the model shows how argumentation expands the sharing of knowledge between dialogue participants. For an individual, critical dialogue has a function of eliciting knowledge that may not otherwise have been used in the current problem. The features elicited in critical dialogue are represented by columns in the mental model tables. Each feature is a dimension along which possible states of affairs can vary. Thus, each new feature increases the number of logically possible situations, i.e., the combinations of truth and falsity. For example, since there are six features (columns) in the final step of our example, there are actually 2x2x2x2x2x2 = 64 possible states of affairs! Clearly, it would be impossible for humans to keep that many possibilities 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 models that needs to be considered is much lower, and does not necessarily increase much at all as new dimensions are introduced. The objective of the proponent 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. In addition, concessions function as assumptions which reduce the range of alternatives to be considered. This example required explicit representation of at most three of the 64 logically possible mental models at any given time.

The objective of the opponent, of course, is to increase the number of mental models, i.e., to force the proponent to consider and respond to alternative possibilities in which her conclusion is not the case. The evolution of mental models for both the proponent and opponent thus provides a vivid record of the progress of a critical dialogue, and clarifies the kinds of moves that each side should make in order to persuade the other.

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 both mental model theory and 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 more 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. Two features clinch its internal status: First, the norms are applied only to facts that are *known* to one or both of the participants. 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. All of these issues require judgments of external reliability.

Dialogue theory does not address the rationale 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, with appropriate norms,

might make sense when one participant has significantly more knowledge and experience than the other; but an information seeking dialogue, with different norms, should be used when one party merely has information that the other party lacks.

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 information to make a decision, but might be insufficient to justify a conclusion when more information or more time is available or the stakes are more serious.

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. This requires judgments about the relative reliability of different belief formation processes as well as the coherence of the alternative mental models with a large store of background information.

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 elements of *externalist* models, which take into account likely outcomes and their associated impact on objectives.

To help dialogue theory bridge the gap between internal and external concerns, it is necessary to provide a third role, that of a *judge*, in addition to those of proponent and opponent (Figure 6). 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.

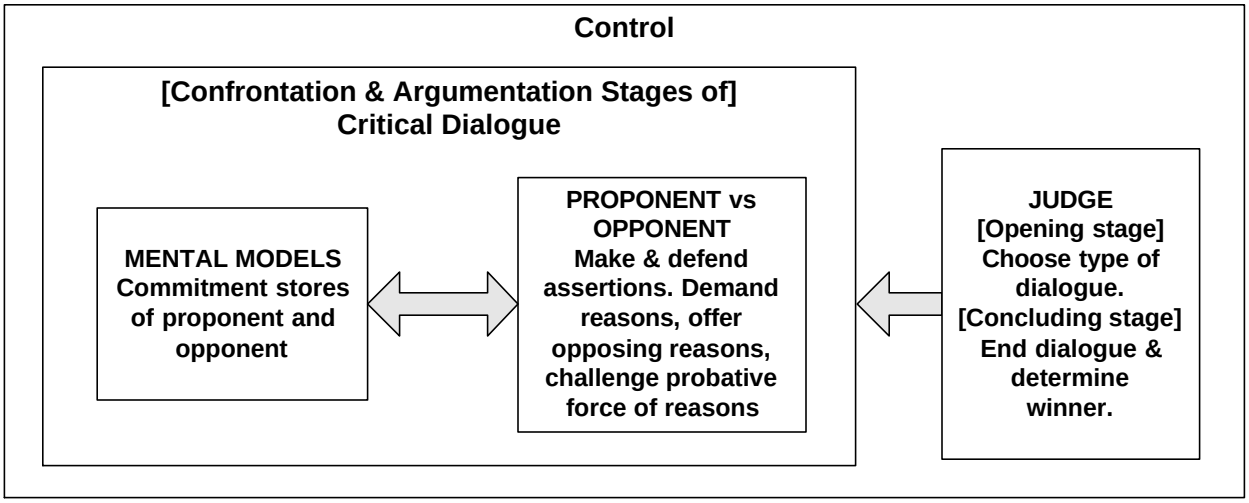

Figure 6. Three part model of critical thinking in terms of stages and roles in a critical dialogue.

Figure 6 shows that each component of our critical thinking model (Figure 1) 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; and the judge determines the overall reliability of the process and regulates it accordingly. The judge is subject to the same capacity limitations as the proponent and opponent, of course, and 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 select and regulate belief forming strategies based on relatively automatic processes shaped by experience, 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.

The introduction of a reliability-based judge has another advantage. It 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 in terms of their reliability, even though they do not themselves involve reason-giving and critiquing. Thus, we can think of the judge as evaluating not only the reliability of different dialogue types, but more generally, evaluating the effectiveness and efficiency of alternative cognitive faculties, and decision making and problem solving strategies. In some situations, taking time to reason may not be the best solution.

Conclusions: What about critical thinking in the Army?

It is appropriate now to summarize some of the implications of this theory for the challenges we laid down at the beginning. 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 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, 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 sense. It not only evaluates possibilities, it generates *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, 1998, p.121).

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, at al., 2000a). 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.

REFERENCES

Berry, D.C. & Dienes, Z. (1993). Implicit learning. Mahwah NJ: Lawrence Erlbaum Associated, Inc..

BonJour, L. (1985). The structure of empirical knowledge. Cambridge, MA: Harvard University Press.

Brandom, R. (1996). Articulating reasons: An introduction to inferentialism. Cambridge MA: Harvard University Press.

Cherniak, C. (1986). Minimal rationality. Cambridge MA: MIT Press.

Chisholm, R. (1977). Theory of knowledge. New York: Prentice-Hall.

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.S., Salas, E. & Riedel, S. (2001). What is critical thinking? Challenge, possibility, and purpose. Arlington, VA: Cognitive Technologies, Inc.

Cohen, M., Thompson, B.B., Adelman, L., Bresnick, T.A., Shastri, L., & Riedel, S. (2000a). Training critical thinking for the battlefield. Volume II: Training system and evaluation. Arlington, VA: Cognitive Technologies, Inc.

Cohen, M.S., Thompson, B.T., Shastri, L., Salas, E., Freeman, J. & Adelman, L. (2000b). Modeling and simulation of decision making under uncertainty. Arlington: Cognitive Technologies, Inc..

Conee, E. & Feldman, R. (2000). The generality problem for reliabilism. In E. Sosa & J. Kim (Eds.), Epistemology: An anthology. Oxford, UK: Blackwell.

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.

Dretske, F.I. (1983). Knowledge and the flow of information. Cambridge MA: MIT Press.

Everitt, N. & Fisher, A. (1995). Modern Epistemology: A new introduction. Cambridge, MA: McGraw-Hill.

Foley, R. (2000). Skepticism and rationality. In E. Sosa & J. Kim (Eds.) Epistemology: An anthology. Oxford, UK: Blackwell.

Goldman, A.I. (1986). Epistemology and cognition. Cambridge, MA: Harvard University Press.

Goldman, A.I. (1992). Liaisons: Philosophy meets the cognitive and social sciences. Cambridge, MA: MIT Press.

Haack, S. (1993). Evidence and inquiry: Towards reconstruction in epistemology. Oxford, UK: Blackwell.

Hamblin, C.H. (1970). Fallacies. Newport News VA: Vale Press.

Harman, G. (1986). Change in view. Cambridge, MA: The MIT Press.

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-Laird, P.N. (1983). Mental models. Cambridge, MA: Harvard University Press.

Johnson-Laird, P.N. & Byrne, R.M. (1991). Deduction. Mahwah, NJ: Lawrence Erlbaum Associates.

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.

Klein, P. (2000). A proposed definition of propositional knowledge. In E. Sosa & J. Kim (Eds.) Epistemology: An anthology. Oxford, UK: Blackwell.

Kornblith, H. (1989). The unattainability of coherence. In Bender, J. W. The current state of the coherence theory. Dordecht-Holland: Kluwer Academic Publishers.

Lehrer, K. (2000). Theory of knowledge. Boulder, CO: Westview Press.

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.

Patel, V. L. & J., G. G. (1991). The general and specific nature of medical expertise: a critical look. In K. A. Ericsson & J. Smith (Eds.) Toward a general theory of expertise. Cambridge UK: Cambridge University Press.

Plantinga, A. (1993). Warrant and proper function; Warrant: The current debate. NY: Oxford University Press.

Pollock, J.L. & Cruz, J. (1999). Contemporary theories of knowledge. Lanham, MD: Rowman & Littlefield.

Quine, W.V. & Ullian, J.S. (1970). The web of belief. NY: Random House.

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.

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.

Sosa, E. (1991). Knowledge in perspective: Selected essays in epistemology. New York: Cambridge University Press.

Sternberg, R.J. & Horvath, J., A. (1999). Tacit knowledge in professional practice. Mahwah NJ: Lawrence Erlbaum Associated, Inc..

Thagard, P. (2000). Coherence in thought and action. Cambridge MA: MIT Press.

Unger, P. (2000). An argument for skepticism. In E. Sosa & J. Kim (Eds.) Epistemology: An anthology. Oxford, UK: Blackwell.

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.

Walton, D.N. (1996). Argumentation schemes for presumptive reasoning. Mahwah, NJ: Lawrence Erlbaum Associated, Inc.

Walton, D.N. (1998). The new dialectic: Conversational contexts of argument. Toronto: University of Toronto Press.

Walton, D.N. & Krabbe, E.C.W. (1995). Commitment in dialogue: Basic concepts of interpersonal reasoning. Albany: State University of New York Press.