EDUCATIONAL PSYCHOLOGIST, *46*(2), 84–106, 2011

ISSN: 0046-1520 print / 1532-6985 online
DOI: 10.1080/00461520.2011.558816

Argumentation, Dialogue Theory, and Probability Modeling: Alternative Frameworks for Argumentation Research in Education

E. Michael Nussbaum
Department of Educational Psychology
University of Nevada, Las Vegas

Toulmin's model of argumentation, developed in 1958, has guided much argumentation research in education. However, argumentation theory in philosophy and cognitive science has advanced considerably since 1958. There are currently several alternative frameworks of argumentation that can be useful for both research and practice in education. These frameworks include Walton's dialogue theory and Bayesian models of everyday arguments. This article reviews and evaluates these frameworks and shows how each can be applied instructionally (e.g., through the teaching of critical questions or probability modeling) and, from a research standpoint, in evaluating the content and quality of informal arguments. It is concluded that attention to these and other contemporary argumentation frameworks can help move the study of argumentation in education forward.

Over the last 15 years, there has been growing attention among educational and developmental psychologists to "argumentation" as a process that may foster conceptual understanding and critical thinking. Beginning with the seminal work of David Perkins (1985), Deanna Kuhn (1991), and James Voss (Voss, Tyler, & Yengo, 1983), this topic area has now blossomed with a variety of books (Andriessen, Baker, & Suthers, 2003; Andriessen & Coirier, 1999; Erduran & Jimenez-Aleixandre, 2008; Kirschner, Buckingham Shum, & Carr, 2003; Muller Mirza & Perret-Clermont, 2009; Reed & Norman, 2004) and special editions of journals (Nussbaum, 2008a; Voss, 2001) dedicated to the topic.

Argumentation research in education has been guided in large part by the Toulmin (1958) model, which is more than 50 years old. Since that time, several other approaches to argumentation have emerged in philosophy and cognitive/computer science, but—with some exceptions—these alternative approaches have not informed educational research in argumentation and may be largely unknown to many readers. This article reviews two of these approaches: (a) Walton's dialogue theory, and (b) the Bayesian approach to argumentation. Although on the surface the two approaches appear vastly different, I argue that in fact the two approaches somewhat complement one another. The article also discusses the implication of each approach for instruction and the assessment of argument quality. It is concluded that attention to these and other contemporary argumentation frameworks can help move the study of argumentation in education forward.

Correspondence should be addressed to E. Michael Nussbaum, Department of Educational Psychology, University of Nevada, Las Vegas, Box 453003, Las Vegas, NV 89154-3003. E-mail: Nussbaum@unlv.nevada.edu

BACKGROUND

What Is Argumentation?

To be clear up front, the term *argumentation* is not used to refer to a debate, although that is one form of argumentation. Rather, it is a process of thinking and social interaction in which individuals construct and critique arguments (Golanics & Nussbaum, 2008). O'Keefe (1982) distinguished between two senses of the word argument, argument-1 ("argument as product"), which consists of a series of propositions in which a conclusion is inferred from premises, and argument-2 ("argument as process"), which refers to the social processes in which two or more individuals engage in a dialogue where arguments are constructed and critiqued. Thus a classroom discussion, in which students are making and evaluating one another's arguments, would be a form of argument-2. This is often called *dialectical argumentation* because a thesis and supporting argument is contradicted with an antithesis (counterargument, refutation, etc.). Theorists have also distinguished between adversarial argumentation, in which the goal is to "win" an argument, as in a debate, and collaborative argumentation, in which students work cooperatively to explore and critique different ideas (Brown & Renshaw, 2000; Gilbert, 1997).

Argumentation provides psychologists with a useful perspective for evaluating students' reasoning. The arguments constructed when students reason—whether expressed in dialogue, a written essay, or some other form—can be judged in various ways (e.g., as shallow or deep, balanced or unbalanced, supported or speculative, simple or complex, etc.). There is a growing body of literature showing that the quality of students' arguments is correlated with learning and conceptual change in science (Alexopoulou & Driver, 1996; Asterhan & Schwarz, 2007; Baker, 2009), understanding of mathematical concepts (Lampert, Rittenhouse, & Crumbaugh, 1996; Schwarz, Neuman, & Biezuner, 2000), better mathematical problem solving (Vye et al., 1997), and comprehension of issues in history and social studies (De La Paz, 2005; Wiley & Voss, 1999). Wegerif, Mercer, and Dawes (1999) have found that engaging in a collaborative form of argumentation known as "exploratory talk" resulted in higher individual scores on a test of fluid intelligence, the Raven's Progressive Matrices. Argumentation can also be motivating and engaging for students (Chinn, 2006). For reviews of these topics, see Chinn (2006) and Nussbaum (2008a).

The Toulmin Argumentation Model

Models of argumentation serve several purposes. One purpose is analytical; models help researchers breakdown arguments into components and study how those components relate to one another. In other words, models help reveal the structure of arguments. A second purpose is normative; models can be used to judge the strength and quality of particular arguments or argument components. Models can also be used to prescribe how an argument should unfold in a dialogue, or to determine appropriate versus fallacious argumentation moves. A third purpose is descriptive; argumentation models can be used to make descriptive and explanatory claims about how people actually tend to argue. As such, argumentation models can be used as a basis for building psychological theories of how individuals comprehend, produce, and evaluate argument information. A specific model may not serve all three purposes just outlined, and it is important to delineate what functions a particular model serves well and what functions it is not intended to address.

The dominant theoretical framework that has been applied to argumentation by educational researchers has been a philosophic model developed by Stephen Toulmin (see, e.g., Bell & Linn, 2000; Chambliss & Murphy, 2002; Erduran, Simon, & Osborne, 2004; Kelly, Drunker, & Chen, 1998; McNeil & Krajcik, 2009; Toth, Suthers, & Lesgold, 2002; Voss et al., 1983; Weinberger, Stegmann, & Fischer, 2005). Toulmin (1958) first presented the model in his book *The Uses of Argument*, where he proposed that the components of an argument could be grouped into one of six categories: (a) claim, (b) grounds, (c) warrants (linking grounds to claims), (d) backing for warrants, (e) rebuttals, and (f) modal qualifiers (e.g., "most," "probably," "possibly"). His classic example concerned a hypothetical lad by the name of Harry.

1. Harry is likely a British subject (claim), because:
2. he was born in Bermuda (grounds), and
3. people born in Bermuda generally are British subjects (warrant),
4. on account of the following British statutes (backing for warrant),
5. unless he became a naturalized American or both his parents were aliens (rebuttals).
6. So the conclusion that he is a British subject is only likely (modal qualifier) and is not certain. The conclusion is subject to exception.

Toulmin proposed this framework as part of a broadside attack on formal logic as a foundation for argument. With the advent of analytic philosophy in the 20th century, formal logic was advanced as a tool to solve (or actually "dissolve") most philosophic problems, which were viewed mostly as conceptual and linguistic confusions. Formal logic uses validity as a standard for judging arguments, which requires the conclusion to follow from premises with absolute certainty (e.g., "All men are mortal, Socrates is a man, and therefore he is mortal"). With the exception of mathematics, however, arguments in most domains are not deductively valid but involve some degree of uncertainty, as was the case with the Harry argument. Toulmin also gave the example of the argument (which I have slightly modified), "Petersen is almost certainly a Lutheran" because "Petersen is a Swede" and "Most Swedes are Lutheran." This syllogism is valid, but how does one know that it is sound (i.e., has true premises)? The standards of analytic philosophy (validity) would require one to back up the universal premise from observing all Swedes, including Petersen. However, if one initially observes that Petersen is a Lutheran, and this information is part of the premises, then there is no point in inferring that information. Toulmin concluded that analytic standards were circular and not how ordinary arguments work. Toulmin claimed that the nature of backing for arguments differed for different domains; an aesthetic argument would use different sorts of evidence for backing than a scientific one, for example. The strength of arguments would have to be judged by domain specific standards. However, Toulmin did make one important generalization: In most cases, conclusions of arguments are tentative and open to exception (hence the importance of the rebuttal and qualifier components).

Evaluation of the Toulmin Model

Interestingly, Toulmin did not initially intend to propose a model of argument. Toulmin (2003) began the preface of the second edition of *The Uses of Argument* by stating, "In no way had I set out to expound a theory of rhetoric or argumentation; my concern was with twentieth-century epistemology" (p. vii). It seems that he did not consider his model that definitive but rather something that was proposed to make his larger philosophic argument.

Despite his original intent, others have certainly received it as a definitive model, and Toulmin (2003) in fact noted that "it would now be churlish of me to disown the notion 'of the Toulmin model"' (p. viii). The model became popularized by textbooks in speech communication (Ehninger & Brockriede, 1961) and English composition (Hairston, 1981; Rottenberg, 1985) and has since been used for a variety of purposes.

One major misconception that has grown about the Toulmin model is that it posits that all arguments have the six components. Toulmin was not, however, making a claim that all arguments contain these components, for he asserted that some argument components may be absent or left implicit; in fact, warrants typically are left implicit unless extreme clarity is needed (Toulmin, 1958). The model is therefore not descriptive of actual social and psychological practices, nor is the Toulmin model primarily normative; he made no claims that a good argument would necessarily have all six components. An individual might provide grounds, or backing, or some other argument component, but the quality of grounds, backing, and so on (i.e., content), is under the model evaluated by domain-specific standards regarding, for example, as to what counts as evidence and the weight that should be given to any piece of evidence. The Toulmin model itself does not provide a basis for judging argument strength. As noted in Toulmin, Rieke, and Janik (1984), although the modal qualifier portion of the argument might describe the strength of the claim (as certain, possible, or probable), that is not the same as evaluating whether the modal qualifier is appropriate (again, domain-specific standards are needed). It is possible to tie some normative criteria—justified on other grounds—to the Toulmin model; for example, arguments with grounds are stronger than arguments without grounds. (The latter are not technically arguments, because an argument is defined as a statement supported by at least one reason [Angell, 1964], but sometimes students and others generate circular or pseudoarguments.) Writing instructors and researchers have also used the Toulmin model to teach students to make their warrants more explicit, which is considered a feature of good writing (Bereiter & Scardamalia, 1987; Fulkerson, 1996b).

These applications of the Toulmin model have allowed it to be used somewhat productively in educational research (see Chambliss & Murphy, 2002; McNeil & Krajcik, 2009), but an overreliance on the Toulmin model among researchers may also be stymieing further progress in the field because the model is not primarily evaluative. As a result, many educational researchers have been left struggling with how best to evaluate the quality of students' arguments (Duschl, 2008; Erduran, 2008). This is a problem even for researchers who do not use the Toulmin model and simply count numbers of arguments and counterarguments, because some arguments or counterarguments can be stronger than others. (See van Eemeren, Grootendorst, Jackson, & Jacobs, 1993, for a discussion of how the notion of counterarguments is underdeveloped in the Toulmin model.)

According to van Eemeren et al. (1996), the Toulmin model is primarily analytical, that is, used to determine the structure of an argument. Only after the parts and structure are specified should an analyst evaluate whether the argument is strong, fallacious, and so on.

Nevertheless, even as an analytic tool, the Toulmin model is somewhat limited. Many have found it difficult to reliably distinguish certain categories, such as warrants and backing (Erduran, 2008), and there are some argument structures found in real arguments, such as convergent arguments (arguments with a number of separate reasons) that do not fit the Toulmin model (van Eemeren et al., 1996).

Because the model is primarily analytical and not a descriptive account of how people argue, it becomes harder to use the Toulmin model to build psychological models of how individuals process and produce arguments. Much current psychological theorizing regarding argumentation relates to why individuals often fail to consider counterarguments, with explanations relating to limitations in knowledge (students may lack knowledge of possible counterarguments; McNeil & Krajcik, 2009), metacognition (students may fail to distinguish and coordinate theory and evidence; D. Kuhn, 1991), interference (the activation of my-side arguments may interfere with efforts to construct counterarguments; Hochs, 1984), or lack of task understanding (students may not see the value of considering counterarguments; Nussbaum, 2005). There is also theorizing as to when and why argumentation may help learning (Nussbaum, 2008a), due to cognitive conflict (Limón, 2001) or knowledge integration (Bell & Linn, 2000). However, none of these various psychological constructs are rooted, at least very strongly, in the Toulmin model.

The Toulmin model has been foundational in establishing that arguments are open to exceptions, unfold in dialectical question–answer exchanges (e.g., grounds are often produced in response to the question, "What have you got to go on?"), and based on various sorts of backing. There have, however, been numerous advances made in philosophy and cognitive science since Toulmin first published his book more than 50 years ago. It is time to begin assessing which of these developments may be useful for argumentation researchers in education.

In this article, I first review Walton's dialogue theory as a more comprehensive framework for analyzing and evaluating arguments. Dialogue theory reflects many of the advances made in philosophy regarding argumentation. The framework takes into consideration various types of argumentative discourse and the participants' goals, various types of argumentation schemes and related critical questions, and how arguments establish a certain degree of plausibility. The second part of the article presents a rather different approach for analyzing and evaluating arguments, based on Bayesian probability theory. Bayesian probability theory, although well established, is just beginning to be applied to dialectical argumentation, and I present several examples of how this could be done. Bayesian analysis provides for a way to gauge the extent that one argument diminishes another, and a systematic way for analyzing arguments that can provide useful insights to participants or researchers. The article ends by discussing how greater attention to these and other contemporary frameworks (and alternatives to the Toulmin model) can help move the study of argumentation in education forward. The implications for both instruction and assessment are discussed throughout the article.

WALTON'S DIALOGUE THEORY

Like the other approaches that are examined here, dialogue theory recognizes that arguments unfold in the dialectical interchange between two or more parties. The term *dialectic* refers to responding to arguments with counterarguments, refutations, elaborations, questions, and other argument-related speech acts to achieve a common purpose among the participants (Walton, 2007). According to Walton (2000), although this approach dates back to the ancient Greeks, it reappeared in analytic philosophy in the 1970s and has been usefully applied to the study of informal fallacies. A central concept of the dialectical approach to argumentation is that of *defeasibility*, that arguments can be defeated by other arguments and these in turn can be defeated. An argument is warranted if it is ultimately undefeated (Pollock, 1987). Defeasibility is a powerful concept because it provides a foundation for warranting arguments.

Although there are a number of philosophers and cognitive scientists (especially in artificial intelligence) that have developed defeasible and dialectical logics (see Atkinson, Bench-Capon, & McBurney, 2006; Carbogim, Robertson, & Lee, 2000), I focus here on one representative approach, that of philosopher Douglas N. Walton. In a prolific series of 40 books over the last 30 years, Walton has synthesized many of the other ideas in the argumentation field and, as such, has one of the most well-developed frameworks in argumentation.

Walton's Framework

Discourse types. There are several levels to Walton's framework, which I have tried to synthesize in Figure 1. Walton's framework is pragmatic, that is, the structure and content of argument is shaped by the joint goals of a reasoner and the other parties with whom she is reasoning. As a result, there are different types of argument dialogues

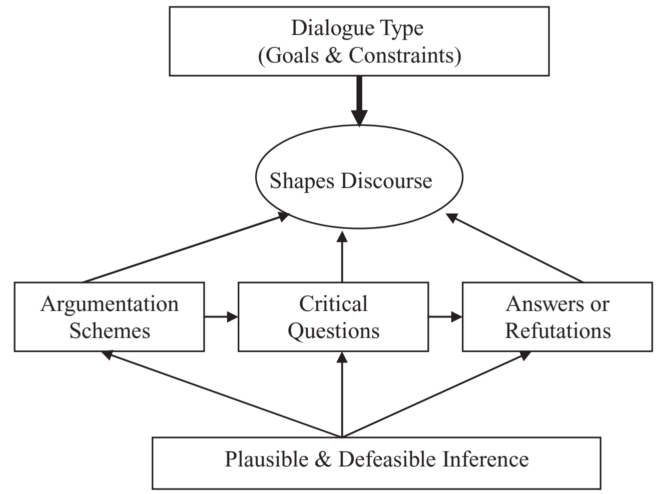

FIGURE 1 Walton's argumentation framework. *Note.* Argument schemes reflect the content and structure of specific arguments; one or more schemes may be used in a dialogue. Associated with each scheme are critical questions that should be discussed about the soundness of each argument. Discourse is also shaped by the goals and constraints of the type of argumentation in which the parties are engaged (e.g., persuasive dialogue, negotiation; Walton, 1998). The argument schemes reflect the microstructure of specific arguments, whereas dialogue type operates at a more general or macrolevel. The macrostructure of argumentation (how various arguments, counter-arguments, and refutations are related to one another) is associated with the "shapes discourse" node.

(Walton, 1998); each type of dialogue has specific goals and is governed by specific rules and constraints that serve those goals. For example, in a persuasive dialogue, the goal is for one party ("the proponent") to persuade another party ("the respondent") of the truth of a proposition, using propositions that the respondent has committed to as premises. The primary constraints are that the parties must make relevant arguments using acceptable argumentation schemes (described next) and—when contradictory commitments are pointed out by the other party—should retract one or more of those commitments. In contrast, in an *inquiry* dialogue, the goal is to establish the truth or falsity of a particular proposition (or that the evidence is inconclusive) by proceeding from premises that everyone agrees are true. The main constraint is that retractions are to be avoided. Unlike persuasive dialogues, inquiries are collaborative and proceed in a cumulative fashion. Another form of dialogue described by Walton is the *eristic quarrel*, where the goal is to emotionally attack the other party so as, for example, to air grievances and to experience emotional catharsis. The main constraints are that the parties should take turns speaking and should not concede defeat, which contrasts with persuasive dialogue where a party should concede defeat if logically warranted. (See Walton, 1998, for an explanation of the other forms of dialogue: negotiations, deliberations, information seeking, and mixed.) One form of dialogue may be embedded in another, but it is improper and fallacious for one party to thwart the goals of the dialogue by subtly and deceptively shifting to another form (e.g., when one party in a critical discussion starts to respond with personal eristic attacks; Walton, 1998).

As intimated by the prior paragraph, the notion of *commitment* is a central aspect of Walton's framework. One often thinks of the goal of arguers as attempting to change a belief of the other party, but *belief* is a psychological construct that is hard to assess (Walton, 2007). Certainly educational psychologists measure beliefs with self-report surveys, but these are not typically used by those engaging in a heated argument. Furthermore, the reports may not be sincere or reflect stable belief (Walton, 2010). Commitments, on the other hand, are more stable and observable. What people express during arguments are propositions, and asserting something is true commits ones to defend the statement unless one withdraws the commitment.

Argument schemes and critical questions. As shown in Figure 1, dialogue type operates at the top or global level of discourse because it frames the general context in which discourse takes place. Dialogue type (goals and constraints) shape argumentative discourse, but the discourse is also shaped by the specific *argumentation schemes* that the parties use in their discourse. An argumentation scheme is a particular pattern of reasoning (i.e., type of argument) that is commonly acceptable. From his analysis of informal arguments provided in speeches, newspaper editorials, and other sources, Walton (1996) identified several dozen schemes, each reflecting a common type of argument (e.g., argument from consequence, argument from sign, argument from expert opinion). Each scheme makes use of a different type of general principle (or warrant, to use Toulmin's terminology). In 1996, Walton published *Argumentation Schemes for Presumptive Reasoning*, which identified 24 schemes, or patterns of reasoning, based on his previous work (these are summarized in Table 1). In a subsequent book, the number of schemes was expanded to 60 (Walton, Reed, & Macagno, 2008). Although the notion of argument schemes can be traced back to antiquity (e.g., in the *topio* of Aristotle), the contemporary notion of argumentation schemes were first articulated by Perelman and Olbrechts-Tyteca (1958/1969), and although there are currently several other typologies of schemes besides Walton's (Grennan, 1997; Hastings, 1963; Kienpointner, 1992), Walton's is the most systematic.

By way of illustration, consider one particular argumentation scheme in-depth, argument from expert opinion. During an argument, it is often appropriate to cite expert opinion. For example, one might argue that people should be encouraged to purchase hybrid cars because such cars produce less carbon dioxide, and most expert scientists believe excessive levels of carbon dioxide are causing climate change. The basic logic of this scheme is as follows (Walton, 1996):

E is an expert in domain D.
E asserts that A is known to be true.
A is within D.
Therefore, A may (plausibly) be taken to be true.

There is also an implicit premise here, specifically that the things that experts say tend to be true. All arguments have some sort of conditional premise, which correspond to the notion of *warrants* in the Toulmin model, which shows how Walton's model is partially based on Toulmin's. In fact, attempts to classify warrants into various types (Ehninger, 1974; Ehninger & Brockriede, 1961) contributed to the development of argumentation scheme theory, as have more recent developments in formal nonmonotonic logics (which are not part of the Toulmin model).

Another important aspect of Walton's framework is the notion of *critical questions* (Hastings, 1963). Associated with each argumentation scheme is a set of critical questions that should be asked by a respondent to determine if a scheme is being appropriately used. Table 1 presents the critical questions for each scheme. So, for example, in the case of expert opinion, one should ask whether the expert truly is an expert in the actual domain in question, whether the expert is unbiased, and so on. Each scheme creates a presumption that the conclusion is true but that presumption can be overturned if a critical question is not satisfactorily answered (Walton, 1996). The notion of presumption is borrowed from jurisprudence, where, for example, a defendant is presumed to be innocent unless the evidence proves her guilty beyond a reasonable doubt. "Reasonable doubt" is a *burden of proof* that must be met to overturn the presumption. In an argument, who holds the burden of proof (proponent or respondent) will vary depending on the critical question involved (Walton et al., 2008). Once a critical question is raised by one of the parties, the other must answer it sufficiently, as determined by the burden of proof (e.g., by a preponderance of evidence). Figure 1 therefore shows that critical questions normally lead to answers. These questions can also lead to refutations of arguments if the question is not satisfactorily answered. So, for example, a party could refute an expertise argument by claiming that the expert is biased or that her credentials have not been established. Refutations, however, may themselves be refuted, as these inferences are defeasible.

As another example of an argument scheme, consider *argument from example* (described near the top of Table 1). Students often back up their arguments with personal anecdotes. For example, in some of my prior work (e.g., Nussbaum, 2005), undergraduate preservice teachers argued about whether they thought watching violent TV caused children to become more violent. One student supported her thesis that TV viewing does cause violence by noting that when she was babysitting for her 8-year-old nephew, he became more violent after watching the TV show *Power Rangers*. This is a specific example that is used to support a conclusion. Critical questions that should be asked about this argument include the following: (a) Is it actually true that the nephew became more violent? (b) How typical is the behavior? (c) How strong

TABLE 1
Walton's Argumentation Schemes and Critical Questions

Scheme	*Critical Questions*
Argument from sign. Observation is evidence of existence of an event or property.	1. What is the strength of the correlation of the sign with the event signified? 2. Are there other events that would more reliably account for the sign?
Argument from example. An example is used to support a generalization.	1. Is the proposition presented by the example in fact true? 2. Does the example support the general claim it is supposed to be an instance of? 3. Is the example typical? 4. How strong is the generalization? 5. Are there special circumstances in the example that impair its generalizability?
Argument from verbal classification. A case ("a") has a property ("F") that classifies it in a certain way, and "F" has the property G (e.g., "Your point of view is heresy, therefore it is wrong").	1. Does a definitely have F, or is there room for doubt? 2. D all F have the characteristic G, or is there room for doubt?
Argument from commitment. Proponent claims respondent is committed to a certain position, and therefore should act/believe a certain way (e.g., "You're a communist, so you should be on the side of the union").	1. Is respondent ("a") really committed to position ("A"), and what is the evidence? 2. If the evidence for commitment is indirect or weak, could there be contrary evidence, or a rebuttal that this case is an exception? 3. Does the belief/attitude in the conclusion really follow from A?
Circumstantial argument against the person. Proponent argues that respondent's claim is contrary to her commitment.	1. Does respondent ("a") claim that everyone ought to act in accord with A? 2. What is evidence that a is not committed to A?
Position to know. You should believe A because a asserts it and she is in a position to know.	1. Is a in a position to know that $A$is true/false? 2. Is a an honest (trustworthy, reliable) source? 3. Did a assert that A is true (or false)?
Expert opinion. An expert asserts that A is true in their domain of expertise.	1. Is the expert (E) a genuine expert in the domain (D)? 2. Did E really assert A? 3. Is A relevant to domain D? 4. Is A consistent with what other experts say? 5. Is A consistent with known evidence in D?
Evidence to a hypothesis. If A then B, B is observed, therefore A is true.	1. Is it the case that if A is true then B is true? 2. Has B been observed to be true (or false)? 3. Could there be some reason why B is true, other than it being because A is true?
Correlation to cause. There is a correlation between A and B, therefore A causes B?	1. Is there a large number of instances? 2. Is there a reverse causal relationship? 3. Can a common cause be ruled out? 4. Are there mediating variables? 5. Are changes in B due to how defined?
Cause to effect. A tends to cause B, and A has occurred; therefore B.	1. How likely is B, given A? 2. What's evidence for generalization? 3. Are there counteracting factors?
Consequences. If A is brought about, then good (or bad) consequences will occur. Therefore , do (or do not do) A.	1. How strong is likelihood that consequences will occur? 2. What's the evidence? 3. Are there other consequences of the opposite value to take into account?
Analogy. One case (C_1) is similar to another (C_2). A is true in C_1, therefore it is true in C_2.	1. Are C_1 & C_2 similar in respect cited? 2. Is A true in C_1? 3. Are there differences between C_1 & C_2 that undermine force of similarity? 4. Is there a similar case with A false?
Waste. Don't stop trying to realize A or all your previous efforts will be wasted?	1. Could past efforts still payoff? 2. Is A possible? 3. Does the value of realizing A outweigh the cost of continuing?
Popularity. Nearly everyone accepts A, so presume A is true or right.	1. Are there reasons not to believe or do A (overriding the presumption)?
Ethos. If a is a person of good moral character, then what a contends (A) is more plausible.	1. Is "a" of good moral character? 2. Is a's character relevant? 3. How strong a weight of presumption in favor of A is warranted?
Bias. Arguer ("a") is biased, so likely has not taken evidence on both sides of issue into account.	1. Is the dialogue of the type that requires participants to take both sides into account? 2. What is the evidence that a is biased?
Established rule. If everyone is expected to do A, then you must do so too.	1. Is doing A in fact what the rule states? 2. Does the rule apply to this case? 3. Are other rules involved? 4. Are there reasons for an exception?

(Continued on next page)

TABLE 1
Walton's Argumentation Schemes and Critical Questions (Continued)

Scheme	Critical Questions
Gradualism. If you take a first step (A), you will eventually be caught up in bad consequences. **Types:** –Causal slippery slope. –Precedent slippery slope. –Classification vagueness. –Arbitrary classification. –Verbal slippery slope. –Full slippery slope.	1. Is A what's being proposed? 2. Do any of the casual links in the sequence lack solid evidence? 3. Does the outcome plausibly follow, and is it as bad as suggested?
Need for help. Person y should help person x if x needs help, and y can help without it being too costly for him or her to do so.	1. Would the proposed action A really help x? 2. Is it possible for x to really carry out A? 3. Would there be negative effects of carrying out A that would be too great?

Note. From *Argumentation Schemes for Presumptive Reasoning,* by D. N. Walton, 1996, Mahwah, NJ: Erlbaum; except for argument from need for help, which is from *Argumentation Schemes,* by D. Walton, C. Reed, & F. Macagno, 2008, New York, NY: Cambridge University Press.

is the generalization? (i.e., Did the nephew become much more violent or just slightly more aggressive?), and (d) Are there special circumstances that should be taken into account (e.g., Was there something else that may have been upsetting the child)? Exploration of these critical questions can help serve to generate refutations, for example, that simply repeating Power Ranger moves does not constitute violent behavior.

Plausibility. As shown in Figure 1, in Walton's framework argumentative dialogue rests on a foundation of plausible and defeasible inferences. So what is plausibility? A proposition is plausible if it is reasonable to accept the proposition. There could be evidence supporting the proposition, but even a conjecture can be plausible if it follows from other plausible propositions and does not conflict with known fact.

Following Rescher (1976), Walton speaks about degrees of plausibility; propositions can be mildly, moderately, or highly plausible. This sounds like probability, but it is not. Rescher developed a formal theory of plausibility as an alternative to probability theory and viewed plausibility as an ordinal concept that allows parties to compare propositions (or entire arguments), judging some as more plausible than others. Plausibility is a function of the reliability of the source of the proposition. For example, Walton et al. (2008, p. 229) gave an example where two experts (Dr. Phil and Dr. Wendy) disagree on whether flax oil reduces cholesterol. Although Dr. Phil is not an expert in nutrition, he is still a scientist, and his opinion that flax oil reduces cholesterol has some plausibility. Dr. Wendy, on the other hand, disagrees and because she is an expert in nutrition, she is a more reliable source, and so an argument appealing to her opinion carries more plausibility. (This is a defeasible conclusion, however, and can be overturned by additional evidence, e.g., that she is biased.) For Rescher, a source of a proposition can be a person, news account, or personal observation, but can also be a claim-authorizing principle or our own intellectual resources (e.g., conjectures that are inferred from other plausible propositions).

It will be important to the development of my argument to distinguish plausibility from probability. Plausibility is more ordinal, whereas probability is more quantitative (i.e., measured on an interval scale). Plausibility is significantly different from probability in another way: Opposing propositions can both be plausible, even highly plausible, but both cannot be highly probable. There might be very good reasons to believe that consuming flax seed oil reduces cholesterol and very good reasons to believe that it does not. For example, even though Dr. Wendy might claim that it does not, an equally qualified expert might claim that it does. It makes no sense, though, to claim that flax seed oil probably reduces cholesterol and that it probably does not. This illustrates how probability differs from plausibility in that probability is more constrained, in that the probability of all possible alternatives need to sum to one. Probability statements are ontological claims about the frequency of alternatives, whereas plausibility judgments are epistemological, on whether it is reasonable to believe a certain proposition, and to what degree (Rescher, 1976).

To assess the plausibility of a proposition, one needs to assess the reliability of the source, be it evidence, other propositions, an expert, and so on. In terms of Walton's framework, this can be done using argumentation schemes. Argumentation schemes provide criteria, expressed in the critical questions, for judging reliability (e.g., examining an expert's credentials, as in the flax seed example). As such, argumentation schemes function to enhance the plausibility of a conclusion (through, e.g., evidence or expert opinion). Argumentation schemes increase plausibility by providing more reasons to believe a claim (thus, by definition, making the claim more reasonable), but the reasons can be "undercut" or weakened if a critical question is not satisfactorily answered (Pollock, 1987). Critical questions reflect common ways of defeating a particular type of argument (Blair, 1999).

Moral and aesthetic argumentation. In making the case for the broad applicability of dialogue theory, it must be asked how the framework could be applied to arguments with a moral, aesthetic, or value-based component. It is important for individuals (including students) to understand the nature of moral argumentation, especially when they discuss moral or policy-related questions in the classroom. According to Walton (2003), moral questions do not have one correct answer, often because individuals have different values or make arguments grounded in different moral systems (e.g., utilitarian, deontological, virtue based, religious, etc.). Nevertheless, having individuals argue with one another on moral issues is still valuable because it can produce greater insight and understanding into the issue and one's own position. It might be said that one goal of moral argumentation is to help the participants develop the strongest positions on both sides of an issue.

One might also argue that plausible reasoning, rather than probabilistic reasoning, better characterizes argumentation related to morality and aesthetics, in that in making a strong moral case increases the plausibility of one's position (but not the probability, as the opposing side could also be highly plausible). For example, in aesthetics, interpretive arguments involve understanding a piece from a particular point of view and making a case that the interpretation is legitimate and interesting (Toulmin et al., 1984). Different points of view do not compete with one another; rather, participants in a successful aesthetic discussion can come to recognize what is valuable in one another's point of view. At the same time, the "plausibility" of a point of view can be undermined and defeated by countervailing evidence, so there is still room for dialectical discussion.

In summary, Walton's framework is based on plausible inference and a dialectical (i.e., defeasible) view of reasoning. As shown in Figure 1, plausible and defeasible reasoning occurs when an argumentation scheme is applied and when a critical question is asked or answered. Each of these "moves" in a dialogue game will enhance or diminish the plausibility of one or more propositions. These moves are embedded in a certain type of argumentative discourse (persuasive, deliberative, information seeking, etc.), each with different types of goals and constraints, and, at a more intermediate level, in various types of argumentation schemes (by making or responding to an argument reflecting a scheme).

Applications of Walton's Dialogue Theory to Education

It is the thesis of this article that there are several alternative argumentation frameworks to Toulmin's that can be useful to both teachers and to educational researchers. The instructional implications of Walton's dialogue theory are discussed first, followed by the implications for researchers.

Instructional applications. In regards to instruction, the most obvious application is that students could be taught about the different argumentation schemes and how to recognize their uses. This is in fact now being done in some undergraduate philosophy courses (Rowe, Macagno, Reed, & Walton, 2006), and a software tool, Araucaria (Rowe et al., 2006), has been developed, using argumentation schemes, that students can use when diagramming arguments. (Walton et al., 2008, also discuss several related software, specifically *Rationale* and *Compedium*.) The approach of directly teaching argument schemes is similar to schema training in mathematics and science, where students learn about various types of schemas for solving mathematical word problems or physics problems (Mayer, 1999). In argumentation, students learn how to recognize a scheme and how to respond (i.e., what critical questions to ask). Prior attempts to teach argumentation schemes have either focused primarily on the Toulmin model, with mixed results (Chambliss & Murphy, 2002; Fulkerson, 1996a, 1996b; Yeh, 1998), or on teaching students an argument–counterargument–rebuttal scheme (D. Kuhn & Udell, 2003; Nussbaum & Kardash, 2005; Reznitskaya & Anderson, 2002; Scheuer, Loll, Pinkwart, & McLaren, 2010). Walton's schemes, however, are richer and more specific and focus directly on evaluation criteria, not just on argument structure. As noted by Nussbaum and Kardash (2005), a counterargument or rebuttal can be of poor quality.

How effective is the direct teaching of schemata? The efficacy of this direct approach still needs to be more fully researched, but a meta-analysis by Abrami et al. (2008) suggests that effect sizes in critical thinking instruction, on both skills and thinking dispositions, are related to how explicitly critical thinking goals and objectives are communicated to students. Of course, argumentation skills must be practiced to be learned and refined.

In addition, research by Richard C. Anderson and colleagues (Dong, Anderson, Kim, & Li, 2008) suggests that critical dialogue is important so that students internalize argument schemata as part of their self-talk and self-regulatory processes (for empirical support, see Reznitskaya, Anderson, & Kuo, 2007; Reznitskaya et al., 2001). This point is consistent with Walton's framework, which views the use of argumentation schemes as part of a dialogue. Teachers must make clear the type of dialogue in which they want students to engage (persuasive, deliberative, etc.) and the norms associated with that type of dialogue. Students would also have to learn to monitor and judge productive versus nonproductive dialogue shifts. This approach could be combined with instruction in a few argumentation schemes relevant to that type of dialogue. Instruction in all the 24–60 schemes would likely result in cognitive overload, at least for younger students.

Critical questions and refutational stratagems. A somewhat different instructional approach inspired by Walton was used by Nussbaum and Edwards (in press) in their intervention study of argumentation in three sections of a seventh-grade social studies classroom. In this intervention, less emphasis was placed on scheme instruction and more on critical questions. In groups of about 10, the students

ARGUMENTS	QUESTION	COUNTERARGUMENTS
Yes, the hungry are suffering, and not enough people are donating to food kitchens. How would you like it if you were hungry?—you would want people to help you.	Should we raise taxes to help feed the hungry?	No, families are struggling to make ends meet and cannot afford higher taxes. Also, it's people's own fault that they are poor and hungry.

CRITICAL QUESTIONS:

Look at all the above arguments, and answer each of the following questions?

	Circle Yes or No		Which Argument?
Is there a **creative solution** to any problems raised?	**Yes**	No	Raise taxes on the rich.
Is the creative solution **practical**? (Consider costs.)	Yes	**No**	There's no cost.
For any argument, can you think of any **other likely explanations?** Or **evidence to the contrary?**	**Yes**	No	Dropping out of school. Bills too high. Not earning enough money from job.
Are any of the arguments **unlikely**?	Yes	**No**	
Are any of the arguments **not as important** as others?	**Yes**	No	Taking money from the war to feed the hungry.

INTEGRATE

Using your answers, explain which side you favor and why is the other side weaker? Or do you favor an "in-between" solution? Give reasons.

Arguments is stronger, in my opinion, because the hungry need help. It's not always their fault that they're starving. We need to help them. Also, the rich should be taxed more so we can feed the poor. Sure families may be struggling to meet ends so it should be voluntary for those families, but those with alot of money they don't need should be taxed.

FIGURE 2 Argumentation vee diagram (AVD). *Note.* From seventh-grade intervention study. Arguments/counterarguments emerged from class discussion the previous day and then were summarized and printed by the researchers onto the form. Students were given 10 min to respond to the critical questions and complete an integrating paragraph. The paragraph instructions encouraged students to reply to arguments on both sides of the issue. AVDs were developed by Nussbaum (2008b) and further developed by Nussbaum, Winsor, Aqui, and Poliquin (2007) and Nussbaum and Edwards (in press). The specific arguments/counterarguments on the above AVD reflect the argumentation schemes *argument from need for help* (see Table 1, and Walton et al., 2008, p. 334) and *rhetorical argument from opposition* (Walton et al., 2008, p. 318).

engaged in weekly oral discussions of current events drawn from *Newsweek*. The instructor facilitated each discussion, asking students to generate and respond to arguments and counterarguments. Before and/or after each discussion, students completed argument vee diagrams (AVDs), such as the one shown in Figure 2, where students had to integrate both sides of the issue to come up with an overall final opinion.

After the first several discussions, I identified some of the argument schemes that were relevant to the type of issues being discussed (e.g., argument from consequences, from an established rule, etc.). I then made a list of the associated critical questions. A lesson was then designed where students were introduced to the concept of critical questions, as tools they could use to determine which side of an issue had

Critical Questions

Critical questions are questions that you can ask about an argument to help evaluate the argument's strength. A strong argument is logically more persuasive than a weaker argument. Here are some examples of critical questions:

Why?

So what?

What's the likelihood?

How important?

Are there reasons for an exception?

How do you know?

Will it solve the whole problem?

Is it practical? (That is, will the solution work?)

How much? How many?

Are there "con" arguments? (These are arguments "against.")

Who will pay for it? What's the cost?

(Are there other critical questions that you can think of?)

FIGURE 3 Handout of critical questions provided to students after 6 weeks of discussion on four different current events. *Note.* The researchers picked schemes and critical questions that seemed applicable to these types of issues and then presented this list to students. Students then completed argumentation vee diagrams (AVDs) and discussed one of these issues. Based on this discussion, critical questions that seemed meaningful and understandable to students were included on a revised AVD that included the critical questions (see Figure 2).

stronger arguments. The critical questions were presented to students as a handout (see Figure 3). The critical questions were from Walton (1996), but I rephrased some to make them more understandable to seventh graders. Pairs of students were then asked to use these questions to complete AVDs on a topic that they had previously discussed, and then discussed this topic further. I wanted to see which of the critical questions seventh-grade students would find understandable and useful. Based on which critical questions were actually used by some of the students, a box of five critical questions were added to the bottom of the AVD shown in Figure 2 (this box was not initially included on the AVDs). I measured over time which of these critical questions—and the evaluation criteria reflected by these questions—were used by students on their AVDs or during oral discussions, compared to a control group that did not received the critical questions box. The dominant argumentation scheme used was argument from consequences (see also Ferretti, Lewis, & Andrews-Wekerly, 2009). There was significant growth over time in asking whether one value is more important than another, for example, whether it is more important to have more family income or to save the planet from climate change. (There was some evidence that some of the students were appropriating a "weighing values" schema, which is a subcomponent of an argument from consequence scheme.) There was also growth in the frequency that students proposed *practical* solutions to problems (e.g., using hybrid cars to reduce greenhouse gas omissions vs. using cars that run on milk). This suggests that students may have been asking themselves the critical question of whether a solution is practical.

This approach focused more on teaching students to ask a few key critical questions and to generate their own critical questions based on the generic ones specified by Walton et al. (2008). In contrast to directly teaching schemes, this approach had the advantage of being easier to fit into a curriculum and involved less cognitive load, because students did not have to learn to recognize specific schemes.

Students also learned specific *stratagems* (Nussbaum & Edwards, in press) with the critical questions. A stratagem is a generic type of brief argument expressed in discourse, such as "I think [POSITION] because [REASON]" or "[CLASSMATE], what if you were in [SCENARIO]?" (R. C. Anderson et al., 2001). R. C. Anderson et al. (2001) showed that elementary and middle school students can effectively learn stratagems from one another during oral discourse, through modeling and appropriation (see also Nussbaum, 2003). Nussbaum and Edwards (in press) took Anderson et al.'s (2001) notion of stratagem one step further and developed the concept of refutational and integrative stratagems. Refutational stratagems are stratagems that specifically refute another claim. Integrative stratagems synthesize arguments and counterarguments. Integrative stratagems may include design claims, which are explanations of how a proposal should be designed that minimize potential problems, and weighing arguments, which are arguments that one value is more important than another, or that some benefit or cost is not as extensive as another. Integrative stratagems are types of refutations because these stratagems serve to reduce the force or importance of an opposing argument.

There are various types of refutational stratagems, and these can be thought of as corresponding to different type of critical questions (recasting the question as an assertion). For example, instead of asking "Is an expert unbiased?" one can claim that "the expert is biased." What is key, from an instructional perspective, is teaching students the evaluative criteria implicit within the critical questions. It is in the teaching of evaluative criteria that this approach goes beyond current approaches that teach students to generate arguments, counterarguments, and rebuttals (e.g., D. Kuhn, 2005).

In this alternative approach that emphasizes the learning of critical questions and refutational/integrative stratagems, argumentation schemes are not neglected altogether; there may be some implicit learning of argumentation schemes. Furthermore, there is some evidence that students may have previous rudimentary knowledge of some schemes (Dushl,

2008) and may mainly need to learn when and where to use the schemes, as well as what critical questions to attach to them.[1] Future research will need to determine whether it is also productive to more explicitly teach the schemes.

Discourse types and shifts. Finally, I propose that it might be useful to teach students about Walton's six types of discourse and about the notion of dialectical shifts. As noted previously, a shift from one type of dialogue to another can be productive if the latter is an embedded dialogue that helps foster the goals of the first. On the other hand, dialectical shifts can be nonproductive if the original discourse goals of the participants are blocked or lost sight of. Classroom discussions are known to easily get off topic and become nonproductive (Brookfield & Preskill, 1999), and it takes a skilled facilitator to move discussions forward so that the original goals are realized. For example, one form of discourse in Walton's typology is a deliberation (choosing between different courses of action). Nussbaum and Edwards (in press) documented how deliberative classroom discussion on the merits of a proposed carbon-tax law shifted to a brainstorming discussion on alternative type of cars (e.g., solar-powered cars). The facilitator used critical questions, however, to move the discussion back to its original deliberative focus. The embedded discourse was useful in helping students understand that some alternative technologies are not yet both cheap and feasible, but some could become more feasible over time. This point was relevant to their deliberations on the merits of a carbon tax. This example shows how Walton's notion of dialectical shifts can be a very useful one for classroom teachers.

Effectively facilitating discussions involves high cognitive load because various arguments and counterarguments have to be attended to concurrently along with assessment of the progress being made toward various dialectical and instructional goals. One question for future research is whether, if students are taught Walton's framework regarding dialectical types and shifts, students can help productively coregulate a discussion with the facilitator, in effect helping to offload from the facilitator some of the cognitive load involved.

Implications for assessment. Having methods for evaluating the quality of students' arguments is important both instructionally and as research tools for documenting the efficacy of various argument-related intervention. However, in educational research on argumentation, there is currently uncertainty on how the quality of arguments should be judged (Erduran, 2008; Scheuer et al., 2010).

Approaches to the argument quality problem. As discussed previously, the extensively used Toulmin model does not provide strong criteria for evaluating arguments. Some argument researchers do not use the Toulmin framework and instead count the number of argument, counterarguments, and rebuttals that students make, or the amount of evidence provided. These are only crude measures of argument quality, however, because argument content is not taken into account. Means and Voss (1996) discussed the features of sound arguments (relevant and acceptable reasons), but the criterion they used for judging a reason as acceptable is unclear, again showing the need for better tools to address the content of arguments. Likewise, Kenyon and Reiser (2006) discussed general evaluative criteria such as having enough supporting data but did not discuss specific critical questions for how to make this determination.

In discussing argument quality, D. Kuhn, Goh, Iordanou, and Shaenfield (2008) suggested that actually refuting the premises of an argument makes for a stronger refutation than simply presenting reason for the opposite conclusion (they called these "Counter-Cs" and "Counter-As," respectively), but it is not clear to me why this should always be the case. A weak Counter-A could be weaker than a strong Counter-C. Somehow again the content of the counterargument or refutation has to be taken into account. In a similar vein, I have suggested that integrated arguments, which address or refute counterarguments, are stronger than unintegrated arguments on grounds of defeasibility (Pollock, 1987); the former are better able to withstand objections and therefore are likely to be stronger. Others have made similar suggestions (see Angell, as cited in Means & Voss, 1996; Erduran et al., 2004; Leitão, 2000). Nevertheless, some integrations and refutations may be stronger than others. In addition, there still needs to be some mechanism for taking the amount of supporting data for respective claims and counterclaims into account.

Recently, Clark and Sampson (2008) surveyed a number of argumentation coding systems that have been used by educational researchers in the areas of science education and computer-supported collaborative learning and then, based on this synthesis, proposed their own approach for how argument quality can be judged. Their approach evaluates arguments on two dimensions: (a) levels of opposition, and (b) conceptual quality. Levels of opposition are scored on a 6-point scale, with the highest score requiring multiple rebuttals with at least one critiquing the grounds of an opposing argument. Conceptual quality is judged based on how the content of a student's argument matches normatively correct scientific ideas. Specific ideas (or *facets*, see Hunt & Minstrell, 1994) are judged as nonnormative (incorrect), transitional

[1]Duschl (2008) coded for several argument schemes during middle school science discussion (e.g., evidence to hypothesis) and found that students exposed to a science intervention focused on developing scientific reasoning used argumentation schemes more frequently than a comparison group. However, the rank order of how often each scheme was used was the same for both groups, suggesting that most students possess many of these schemes in a rudimentary form. According to Duschl (2008), instruction is needed so that students learn the "epistemic criteria" for applying the schemes and judging the strength of arguments. Duschl did not directly discuss the notion of critical questions, but such questions presumably reflect, at least in part, epistemic criteria.

(partly correct), or normative (correct). An example of a non-normative facet would be "all objects have a natural temperature," whereas an example of a normative facet would be "objects in the same room reach the same temperature." A transitional facet would be "all objects in the same room will reach a temperature close to, but not identical to, the surrounding." Conceptual quality is judged on a 4-point scale, with the highest level requiring an argument to contain several normative facets related to both explanation and evidence.

For this approach to work, there needs to be a scientific consensus on what the normatively correct ideas, or facets, are. This may be the case when teaching about thermodynamics or other well-established bodies of knowledge, but this situation is not always the case. For example when students discuss scientific controversies or social issues, experts may disagree on what the normatively correct ideas should be. In these cases, other methods must be used to assess argument quality. Furthermore, even when Clark and Sampson's approach is used, researchers may wish to supplement the approach with other methods, given the multidimensional nature of argument quality. For example, researchers may wish to examine how well organized an argument is, or whether a student's refutations are convincing.

Walton's dialogue theory and assessment. What does dialogue theory contribute to the challenge of assessing argument quality? In my view, the critical question component is the most valuable. First, researchers can determine if students themselves are asking the critical questions or making arguments that reflect the evaluative criteria contained in the questions. I have shown that his can be done reliably (Nussbaum & Edwards, in press) and asking critical questions is a sign that students are reasoning at a sophisticated level. (Evaluative criteria can also be instantiated as refutational statements, and researchers can code for these as well.) Second, researchers can determine what sort of argumentation schemes students are using and can ask how well students are using the schemes as well as addressing the questions associated with a scheme. By way of illustration, if students use an example to support an argument ("argument from example scheme"), researchers should ask, "Is the example typical of the kinds of cases the generalization covers?" (Walton et al., 2008), which is one of the critical questions associated with that scheme. Answering these questions will require some professional and subjective judgment on the part of researchers, but that does not mean such judgments cannot be made reliably after some practice by scorers. For example, writing researchers have been able to make reliable judgments about some aspects of the quality of students' writing (Reynolds & Perin, 2009), including holistic judgments about the persuasiveness of student arguments (Ferretti et al., 2009). Walton's dialogue theory is useful, however, by providing a large range of specific criteria that could be incorporated into scoring rubrics, or used in a qualitative analysis of how student argumentation may improve over time.

The aforementioned approach assumes that researchers can identify the major or predominant argumentation schemes that students are using (or should be using) in their discourse. The evidence is mixed as to whether researchers can reliably determine which argumentation schemes are being used. Ferretti et al. (2009) were able to reliably determine which of seven schemes were used in students' opinion essays; however, Duschl (2008) found—in investigating oral discourse around a science curriculum—that several of the schemes could not be reliably distinguished. Walton et al. (2008) explained that some schemes are subcategories of other schemes and that the exact relationship among schemes is still being determined; this fact might explain some of the difficulties in reliably coding for schemes. However, Nussbaum and Edwards (in press) did not code for schemes at all, just for critical questions and refutations, on the grounds that what is most important is whether students are evaluating one another's arguments.

Although this article has focused on argumentation schemes and critical questions, Walton's dialogue theory is broader than this, and makes reference to such constructs as dialogue type, dialectical shifts, and plausible reasoning. Walton (2007) himself presented a list of criteria that researchers should consider in judging argument quality; he recommended that argument discourse be judged as to whether the goals of the discourse were met, what standards of evidence (appropriate for the type of discourse) were met, and the depth of discussion. Walton saw the depth of discussion as a function of six factors: (a) number of arguments brought forth, (b) how many of these arguments were defeated, (c) number of implicit premises revealed in the discussion, (d) how revealing these implicit premises helped participants gain greater insight into the strength and weaknesses of their positions (called the *mautic* effect), (e) how well the discussion was informed by relevant facts, and (f) how strongly argumentation throughout the dialogue "supported or refuted the fundamental thesis at issue" (Walton, 2007, p. 118). Some of these criteria may be more useful to a qualitative rather than quantitative analysis of argument quality, but more research is needed to determine this.

My main point is that Walton's dialogue theory offers researchers in education a useful alternative to the Toulmin model, and a useful extension of an argument–counterargument–rebuttal model. This can help move the field of argumentation research in education forward in several ways. First, much current work in argumentation emphasizes developing classroom norms of discussion (R. C. Anderson et al., 2001; Webb et al., 2008; Wegerif et al., 1999) but does not distinguish among the various types of argumentation that are specified in dialogue theory. There is also little research on appropriate versus inappropriate shifts among the types of dialogue. Although in the literature on scripting computer-supported collaborative learning, there has been some work on shifting discourse from one activity structure to another (Haake & Pfister, 2007; Kopp & Mandl, 2007), for the most part dialectical shifting is an understudied area. Second, there has been little work on combining argumentation approaches with approaches for stimulating higher order questioning (Chin & Osborne, 2010), and the emphasis in dialogue theory on critical questioning would take the field in a new directions, both in terms of instruction and assessment. In particular, although there has been a lot of work performed on how to stimulate counterargumentation (Andriessen, 2006; Nussbaum & Kardash, 2005; D. Kuhn, 2005), critical questioning is another powerful tool that can be taught to students for generating counterargument (Nussbaum & Edwards, in press). Third, dialogue theory has numerous schemes that relate to practical reasoning regarding ill-structured problems; weak applicability to ill-structured problems is another limitation of the Toulmin model (Voss et al., 1983). Finally, as discussed previously, dialogue theory is a tool that would allow researchers to pay greater attention to the content and quality of students' arguments. There are various dimensions to the notion of argument quality, and Walton's framework specifies many of these dimensions in more detail than previous approaches (e.g., Means & Voss, 1996). On the other hand, I do not argue that dialogue theory provides a complete solution to the problem of assessing quality, and researchers should consider other developing approaches to this issue as well. In the next section of the article, I explore a somewhat different approach to argument evaluation based on probability theory.

A BAYESIAN APPROACH TO ARGUMENT EVALUATION

Arguments can be judged as strong or weak along a continuum, with stronger arguments judged to be of better quality. Individuals are often asked to judge whether the arguments and evidence on one side of an issue are stronger than those on the other. But on what basis is one to judge argument strength? Defeasible approaches to argumentation do not really address this issue. When does a refutational stratagem or argument really "defeat" an argument versus just reducing its strength? Furthermore, if the strength of an argument is reduced, at what point is the strength of an argument outweighed by those on the other side? To answer these questions, researchers need a theory of argument strength. Also, because contemporary cognitive theory holds that people represent ideas and the associations among ideas (e.g., arguments) with a certain amount of strength (J. R. Anderson, 1990), a theory of argument strength might inform explanatory psychological accounts of how people process arguments (see, e.g., Kemp & Tenenbaum, 2009).

Some philosophers have held that the strength of an argument is a function of how likely the premises make the conclusion (Copi & Cohen, 1998); stronger reasons make the conclusion more likely. This means that argument strength can be thought of in probabilistic terms. Most arguments do contain a probabilistic component. Many arguments are inductive, which by definition are probabilistic, but even moral, policy, or legal arguments typically have a probabilistic component to the extent that reference is made to probable outcomes or expected utility (Savage, 1972).

The central role of probability in argumentation makes Bayesian probability theory an appealing framework for judging argument strength. The Bayesian school of probability holds that probability is a subjective measure of the strength of belief that can be applied to single events, for example, "It will rain today." The probability is how strongly one believes the proposition to be true (with 1.0 indicating certainty and 0.0 indicating complete denial). Much of Bayesian theory is centered on Bayes' theorem, which provides a mathematically "correct" way of updating prior probability beliefs given new information. (The updated probability belief is known as the posterior probability.) Bayesian theory is well developed in statistics, where it is used, for example, in decision theory, sample size estimation, model selection/averaging, and small sample inferencing (Press, 2003), and has also been applied to the philosophy of science (Howson & Urbach, 1989).

The Bayesian view is applicable to argumentation, because argumentation also involves changing beliefs in propositions through reasons and evidence. It has traditionally been applied to evidence-based arguments, and an example of this is provided in the Appendix along with an explanation of Bayes' theorem. The example involves a hypothetical woman named Mary who initially believes she is not pregnant with a prior probability of 40%. However, after taking a pregnancy test with positive results, her probability estimate should rise to 91% (if she reasons according to Bayes' theorem). The example assumes that the pregnancy test is fairly reliable, specifically, that if she is pregnant, the test will give a positive result 80% of the time and has a false positive rate of only 5%. (These last two figures I call the evidential likelihoods, with the ratio of the two likelihoods indicating the "strength" of the evidence; Royall, 1997).[2]

The Bayesian framework essentially provides rules for how to change probability estimates given new estimates. We see that the framework provides an orderly way of analyzing a set of arguments, making the assumptions explicit and encouraging reflection on those assumptions.

A major issue in Bayesian theory is whether the likelihoods can be estimated objectively. The school of objective Bayesianism holds that the likelihoods as well as the priors should not be set in any old way but should be based on empirical evidence and other rational considerations (Press, 2003; Williamson, 2005); for example, a likelihood could be estimated using random sampling.

[2]Royall (1997) conceived of the evidential likelihood more as the ratio of these two likelihoods when comparing two hypotheses, $\frac{P(H_1)}{P(H_2)}$, which is the odds of H_1 when $P(H_2) = P(1 - H_1)$, that is, when $P(H_1) + P(H_2) = 1$. In other words, when the two hypotheses are exhaustive and mutually exclusive.

On the other hand, the use of random sampling requires certain assumptions that the method that one uses is in fact random. This may be a reasonable assumption to make in many instances, but the assumptions should be open to critical questioning. For example, in the Mary case, to set the false positive rate one might use a study by the manufacturer of Mary's pregnancy test, based on a random sample of people who used the pregnancy test (and other indications of pregnancy). However, how reliable were these other indicators? Was the manufacturer trustworthy or did it make decisions in conducting the study that biased the results? Based on the answer to these critical questions, one might choose to make some adjustments to the manufacturer's estimates, but rational argumentation is needed to show that the magnitude of any adjustment is reasonable.

Some argue that the Bayesian approach is utterly subjective (Bradley, 2003) and therefore cannot produce consensus on probability estimates. Others, such as Walton (2004) argue that Bayesian probability estimates are simply "plausible guesses" that do not rule out other plausible estimates (see also Thagard, 2004) and that it is best to conceive argumentation as ultimately dialectical. Dialectical argumentation can produce consensus, or at least can narrow the scope and range of disagreement, especially when informed by data. Furthermore, Aristotle held that dialectical arguments (between people) can be resolved through appeal to *endoxa* (as cited in Walton et al., 2008), which are beliefs, goals, and values that are held in common by most members of a society. To the extent that certain practices, such as a particular method of random sampling, are accepted as valid, then data collected through that practice can be used to constrain the space of plausible likelihood values.

Dialectical argumentation can therefore strengthen Bayesian analysis, but can Bayesian analysis strengthen dialectical argumentation? I believe that it can, at least some of the time, by helping to yield insight into the strength and weaknesses of many informal arguments, and also by providing a fairly systematic way of analyzing informal arguments. I support this claim with two examples of the analysis of informal arguments drawn from the classroom discourse of middle school students.

Hunger

In a recent study (Nussbaum & Edwards, in press), seventh-grade students engaged in argumentation on whether taxes should be raised to help feed the hungry and homeless. The author facilitated several oral discussions and then analyzed the arguments that students made.

Most students were opposed to the proposal to raise taxes because it would mean additional expenses for their parents and therefore fewer toys for them. In addition, many students argued that students who are homeless or hungry are in that state due "to their own fault," generally because they dropped out of school. I sat down after class to synthesize and evaluate the various arguments made. The last argument could be modeled statistically, using the odds version of Bayes' theorem, as

$$\begin{aligned} &Odds\,(dropout\ \ given\ \ hungry) \\ &\quad = \left(\frac{P(hungry\ given\ dropout)}{P(hungry\ given \sim dropout)}\right) * Odds\,(dropout) \end{aligned}$$

Recall that although the prior and posterior odds are considered subjective, one can still set the likelihood values using somewhat objective methods such as random sampling. From an Internet search, I located a report by the Economic Policy Institute (Boushey, Brocht, Gundersen, & Bernstein, 2001) that provided empirical data from which the likelihoods could be estimated, yielding the following calculation:

$$Odds\,(dropout\ given\ hungry) = \left(\frac{20\%}{20\%}\right) * 0.43 = 0.43$$

If someone were hungry, the odds that they had also dropped out of school were less than 1. So the argument that "it's their own fault that they are hungry because they dropped out of school" was shown to be a weak argument (because it did not make the posterior odds favorable, that is, above 1.0). Most people who drop out of school still find jobs, and more important, the false positive or "counterargument" term, *P(hungry given* $\sim$ *dropout*), was well above zero, suggesting that there are other reasons that people can become hungry (e.g., losing a job). Furthermore, many of the hungry are children and are not in that state because they "dropped out"; it is clearly not "their fault" that they are hungry. I raised this issue with the students during a follow-up discussion, and there seemed to be some attitude shifting among a couple students. The main lesson here is that probability modeling can be used to get a better handle on the issue and help identify major points or counterarguments that should be introduced into the discourse. It could also be used to help a research team gain some sense of the stronger versus weaker arguments that were made. Bayesian probability modeling is one way of organizing an evaluation.

On the other hand, I did not base an entire coding scheme based on the Bayesian framework, in part because of the subjectivity involved (which creates a challenge in achieving interrater reliability). But the subjectivity in the Bayesian approach can be constrained by rational and empirically grounded arguments, enough to make the approach useful in generating insights that can be used instructionally. Also, having to fill in the values of a Bayesian argumentation schema (e.g., Bayes' theorem) forces one to (a) think about an issue from different perspectives, (b) reflect on one's range of certainty and uncertainty, and (c) conduct a simple computer simulation (e.g., calculating the product) that integrates various facets of the issue together and affords further reflection. (One can reflect, e.g., on why the solution comes out a certain way and how the solution is affected by a change in assumptions.) When also drawing on empirical studies, the approach provides a data-based method of reasoned argument that can in part refute arguments based on biased personal experience or political ideology. This is an important instructional objective in developing students who are critical thinkers (Halpern, 1997).

American Torture Policy

The next example is illustrative of some of the challenges in applying a Bayesian analysis to argumentation, and some modifications. The same seventh-grade students who discussed the hunger issue also discussed whether the U.S. government should be permitted to torture suspected terrorists.

The first challenge was to determine whether Bayes' theorem was applicable. The theorem is applicable when empirical evidence is presented; however, none of the students' arguments contained actual empirical evidence. Nevertheless, modeling what I call the *probability structure* of the students' arguments was found to be useful, that is, what probability arguments were being made and how these arguments related to one another. For example, some of their arguments were (a) torture can help extract information to stop another 9/11-type terrorist attack, (b) tortured people often lie just to stop the pain, and (c) innocent people can also be mistakenly suspected as terrorists and tortured. In evaluating these various arguments, the probable benefits of torture depend on torturing a person who actually (a) is a member of a terrorist organization, (b) knows something relevant, and (c) will not lie just to make the pain stop. The probability of any one of these events seems low, and the joint probability (from multiplying the three) must be very low.[3] Therefore, this simple analysis revealed severe problems with the case for using torture.

This is, however, a cold, utilitarian analysis and ignores the role that emotions can play in argumentation in giving proper weight to things one cares about (Thagard, 2006; Walton, 1992). In fact, the students' discussion of torture became emotional at times, with one student, Luis, proclaiming,

> No because who are you to torture someone, who are you . . . God? No. You don't have the power to say, yes, I'm going to torture someone and be happy about it. No, because it is wrong to torture someone.

Another student, Calvin, responded,

> Well, who are *they* to go around planting bombs, threatening other lives, and um, they not getting like at least some type of pain. Because what if those bombs went off before we even knew about it.

These students were quite emotional and passionate about their viewpoints. High emotions can certainly interfere with rational (and probabilistic) thinking, but argumentation completely devoid of emotions also runs the risk of disengaging students and producing poor decisions (Damasio, 1994; see also Nussbaum & Broughton, 2008).

This discussion points to another challenge facing Bayesian analysis: how to apply Bayesian analysis to moral issues (which, of course, are also typically highly emotional ones). The traditional Bayesian approach to analyzing moral issue is to assume a utilitarian moral framework: Pursue the policy that will produce the greatest good for the greatest number (Bentham, 1789/1988) or will maximize happiness (Mill, 1863/1979). Savage (1972) and others have incorporated Bayesian analysis into expected utility theory, where probability analysis is used to determine what sorts of policies are likely to maximize happiness. More specifically, likelihood ratios (i.e., a likelihood divided by a false positive rate) can be incorporated into benefit/cost ratios (reflecting probable benefits over probable costs). Applied to the torture example, some might argue that even though the probability of obtaining useful information through torture is low, the benefits of avoiding another 9/11-type terrorist attack are so high that this outweighs the costs. (On the other hand, to really obtain any benefits at all, one would have to torture lots and lots of people to obtain a "hit," and many of the people tortured would be innocent; not only is this a "wrong" in its own right, but would (and did) damage to American's constitutional values and reputation; Pyle, 2009).

In my view, Bayesian/utilitarian analysis can help students explore the costs and benefits of a particular proposed policy in a systematic fashion. Moreover, it can also be used as a springboard to introduce and discuss other ethical frameworks. For example, in the argument about torture presented earlier, Luis makes a more duty-based (or *deontological*) argument against torture. It would likely be feasible to discuss with these students notions such as the Golden Rule ("Treat others as you would like to be treated") or at higher grade levels, Kant's (1788/1996) categorical imperative. The need

[3]This analysis could be faulted, however, because it assumes that the various component probabilities are statistically independent. If the component probabilities are statistically dependent, meaning that the occurrence of one event makes another more or less likely, then the following formula must be used instead of simply multiplying the probabilities:

$$P(A\&B) = P(A) * P(B \mid A)$$

It is very plausible that an individual is more likely to "know something relevant" if they are a member of a terrorist organization. A more appropriate probability structure would involve the term "probability knows something relevant given that an individual is a member of a terrorist organization," that is, $P(B \mid A)$. The statistical dependency is likely to increase the joint probability somewhat.

Modeling statistical dependencies will make the probability structure more complex. These dependencies can be modeled with graphical methods, specifically causal Bayesian networks (Pearl, 2000), although these networks can, given too many dependencies, become computationally intractable, unless very computationally expensive statistical techniques are used (Glymour, 2001). For analyzing arguments, it is likely not productive to consider all possible statistical dependencies, but if a plausible argument can be made that a particular dependency is strong and important, then an alternative to an independence model should be considered.

to act virtuously (i.e., virtue ethics; Statman, 1997) is another moral framework that could be applied. My point is that Bayesian analysis can be used not only as a springboard to explore costs and benefits but also to explore differing ethical systems. As was noted in the discussion of Walton's framework, the pedagogical goal is not to really resolve the issue through argumentation but to gain greater insight into different points of view, including one's own.

Applications of the Bayesian Argument Framework

In support of the thesis that there are useful alternative frameworks to the Toulmin model, I now examine some of the ways the Bayesian argument framework could be applied by educational researchers, both in terms of instruction and assessment. The most obvious instructional application of this framework is teaching students about probability and how to identify the probability structure of a set of arguments. It would not be far-fetched to have students in a mathematics class conduct Internet research to fill in a contingency table on a social issue, such as hunger (as in Table A2), and encouraged to make some probabilistic judgments. Given the central role of probability in argumentation and decision making, students need much more instruction on probability (Lovett, 2001; Nisbett, 1993), anchored in real-world issues and at earlier ages. Students also need practice in estimating whether probabilities are high or low. This would help students use real numbers in their arguments and to appreciate the value of data in supporting or refuting arguments. The Bayesian framework also affords an opportunity to teach about the nature of scientific modeling and to promote model-based reasoning (Windschitl & Thompson, 2006). More research is needed, however, on the best way to accomplish these objectives.

Probability concepts are especially applicable to physical or social systems that display some natural variability or randomness. Educational researchers have examined and scaffolded student argumentation relative to such systems, for example, related to natural selection (Asterhan & Schwarz, 2007; Reiser et al., 2001), adaptation (Linn, Shear, Bell, & Slotta, 1999), social and policy issues such as capital punishment or nuclear power (D. Kuhn & Udell, 2003; Nussbaum & Edwards, in press; Resnick, Salmon, Zeitz, & Wathen, 2003), and sociotechnical issues such as genetically modified foods (Mason & Scirica, 2006). These efforts, however, have not addressed probabilistic reasoning explicitly. There has been a concern for data analysis, in particular whether different pieces of evidence support or refute alternative hypotheses (see Bell & Linn, 2000), and probability modeling could be used to produce a richer analysis in some domains. There is a variety of argument mapping software (see Kirschner et al., 2003; Scheuer et al., 2010), and some of these approaches could be extended to include probability modeling at more advanced grade levels.

Some might argue that probability concepts are too advanced for some of the grade levels studied (e.g., seventh grade) or that students have a hard-enough time simply focusing on converging lines of evidence (Kelly, Regev, & Prothero, 2008). However, these are claims that need to be tested with empirical evidence (Metz, 1998).

In regards to assessment, probability modeling has some promise as a research tool for evaluating the quality of students' arguments. As noted previously, the strength of many arguments is to a large extent a function of how probable the premises make the conclusion. It may not be practical for researchers to engage in such modeling for every argument generated in a research study, but the major arguments could be identified and some ranked as stronger than others. Then transcripts can be coded for the arguments that students made and whether they made the stronger arguments (see Mason & Scirica, 2006).

Another possible application is for researchers to use the Bayesian framework as an orderly way of assessing a set of arguments (perhaps in conjunction with dialogue theory), so as to gain insight into the issue, and then to feed these insights back to students through instruction, or by "seeding the discourse" (Palincsar, 1998). This was illustrated with the hunger example, where the fact that many of the hungry were children was then fed back to students as a stimulus for further discussion and argumentation. This approach can help students reframe an issue to see it from a different and more productive perspective (Greeno & van de Sande, 2007).

Comparison With Existing Research Trends

Like Walton's dialogue theory, the Bayesian framework is another alternative to the Toulmin model for analyzing informal arguments. In contrast to Walton's dialogue theory, which can potentially take the argumentation field in new directions, the Bayesian framework is more consistent with existing trends in the literature regarding encouraging students to consider alternative viewpoints and hypotheses (D. Kuhn, 2005), providing supporting evidence (Sandoval & Millwood, 2005), and reflecting on one's epistemological warrants (Garcia-Mila & Anderson, 2008). The framework is more powerful than the Toulmin model in providing a set of systematic analysis tools that can be applied to these issues and that can provide researchers and students with greater insight into the strength and weaknesses of various arguments. Bayesian methods also provide a framework for judging argument strength even when there is only a small amount of data available, although according to plausibility theory, small-sample analyses are less plausible (because the source of the likelihood estimates are less reliable).[4]

[4]Subjective and objective Bayesians will likely disagree on the merits of small sample analysis; a lot will depend on the confidence one can place in the likelihoods.

Some researchers have found argumentation more useful for decreasing (rather than increasing) beliefs in theories (Baker, 2009), specifically in the domain of physics. The same might be true with issues such as evolution, climate change, or medical vaccinations, where I fear that the presence of low-quality, public argumentation in the mass media or on the Internet may serve to foster skepticism rather than commitment, based on a (mis)perception that experts disagree and that scientific warrants are thereby weak (Mason & Scirica, 2006; Sinatra, Southerland, McConaughy, & Demastes, 2003). This can be explained by the Bayesian framework because dialectical argumentation involves consideration of more alternative hypotheses, thereby increasing the false positive rate and diminishing the size of the probability estimate. However, one might use the Bayesian framework to predict that the strength of beliefs (particularly in scientific theories) could be increased by having students reflect on and estimate the numerator likelihoods in Bayes' theorem and reflect on the epistemological justification of these estimates. Because it cannot be assumed that people are Bayesian reasoners, this would need to be combined with other conceptual change techniques (Sinatra & Chinn, in press).

The Bayesian approach certainly needs further development, but psychological researchers are making strides in this regard. For example, Hahn and Oaksford (2006, 2007) used Bayesian analysis to better understand why an argument might be fallacious in one context but not in another. Griffiths, Kemp, and Tenenbaum (2008) developed hierarchical Bayesian techniques for analyzing causal inductive and scientific arguments that have both normative and descriptive applications (i.e., explanations of human cognition; see also J. R. Anderson & Lebiere, 1998).

SUMMARY AND DISCUSSION

In this article, it is argued that greater progress in the field of argumentation in education could be made by using alternative frameworks to the Toulmin model. These alternative frameworks can more directly evaluate the content of arguments as opposed to counting argument components. There have been significant advances in argumentation theory since the introduction of the Toulmin model in 1957. There are defeasibility-oriented frameworks, such as Walton's dialogue theory, and more "epistemic" approaches (Williamson, 2005), such as the Bayesian framework, which assess the probability that a conclusion is true. These frameworks identify various important concepts that could be taught to students, such as argumentation schemes, critical questions, refutational and integrative stratagems, dialectical shifts, plausibilistic reasoning, and probabilistic reasoning.

An open question is how the two frameworks relate to one another. Although a full exploration of this question is beyond the scope of this article, the analysis presented here suggests that Bayesian analysis actually depends on dialogue theory, because critical questions need to be asked about, and rational arguments made about, the evidential likelihoods. Conversely, Bayesian analysis adds to dialogue theory a mechanism for assessing "argument strength" so as to better determine when a critical question is satisfactorily answered, or exactly when one argument defeats another. The Bayesian approach cannot completely eliminate the uncertainty and subjectivity involved in assessing arguments, but it can constrain the uncertainty somewhat. Bayesian methods also provide a systematic way of organizing an analysis, which can help generate insights that can be used instructionally.

Another issue is to what extent these frameworks are truly alternatives to the Toulmin model. They are alternatives in the sense that researchers can use them as the dominant theoretical framework for a study rather than the Toulmin model. But from a technical standpoint, they are not inconsistent with the Toulmin model. As noted previously, dialogue theory is partially based on the Toulmin model, and it is shown in the appendix how a Bayesian analysis could be fit into the Toulmin model. However, both frameworks supplement the Toulmin model in very extensive ways. There has been some very good work in argumentation based on the Toulmin model (e.g., Erduran et al., 2004), and, as one anonymous reviewer pointed out, one should not assume that "old is necessarily bad." But old is sometimes limited, and the case is made here that greater progress in the field could be made if one of these alternative frameworks were used to frame more educational studies in argumentation.

I also do not mean to imply that these two alternative frameworks are the only ones that should be considered. There are other argumentation theorists working in the dialectical tradition other than Walton (e.g., Pollock, 1987; van Eemeren et al., 1996; van Eemeren et al., 1993; Vorobej, 2006) as well as in the epistemic tradition other than Bayes (e.g., Grennan, 1997; Parsons, 2001). The field of argumentation in education is young and evolving; progress in the field might be better characterized as incremental and fluid (Lakatos, 1978) rather than as a clash of competing paradigms (T. S. Kuhn, 1970).

As an example of how argumentation frameworks evolve, and of the utility of Walton and Bayes' conceptualizations, consider the framework that the author developed and used in several previous studies: *argument–counterargument integration* (Nussbaum, 2008b; Nussbaum & Schraw, 2007). (This has some similarity to frameworks developed by other researchers, e.g., Erduran et al., 2004; Leitão, 2000; Suedfeld, Tetlock, & Streufert, 1992.) Argument–counterargument integration involves having students integrate arguments on both sides of an issue while generating an overall final conclusion. I have argued that such integrated arguments are stronger than nonintegrated arguments because, by responding to potential counterarguments (accepting or refuting them), such arguments become stronger and more warranted (Pollock, 1987). I developed the argumentation vee diagram shown in Figure 2 as a scaffold to support and facilitate integrative reasoning. The diagram asks students to assess whether arguments on one side of an issue are stronger than those on the other side, but a major shortcoming was that students needed instruction in logical criteria for making this judgment. The AVDs were therefore adapted to add selected critical questions from dialogue theory, and these same questions were emphasized in oral discussions between me and middle school students prior to the completion of the AVDs. As shown in the previously presented examples, Bayesian analysis of the arguments and counterarguments that students generated was also used to help assess their arguments and to better frame key issues for students. This shows that the two approaches can be used in conjunction with one another. I eventually hope to develop a theory of argument–counterargument integration that shows how people tend to weigh and integrate opposing arguments, either naturally or with various types of prompting.

Alternative argumentation frameworks will be critical to this and similar endeavors that seek to provide a richer and more elaborate picture of argumentation processes. Although this article has focused on the normative aspects of dialogue theory and Bayesian analysis, the frameworks also suggest a variety of empirical issues of potential interest to educational psychologists. For example, what is the structure of individuals' argumentation schemata (Reznitskaya & Anderson, 2002)? What is the role of critical questions in fostering understanding (Aulls, 1998; Chin & Osborne, 2010; Graesser, Baggett, & Williams, 1996)? How well do students manage dialectical shifts in discussions and well as in their own thinking (Walton, 1998)? When individuals judge the strength of an argument, to what extent do these judgments reflect probability or expected utility versus some other construct such as explanatory coherence (Thagard, 2000, 2004; Ranney & Schank, 1998)? Last but not least, to what extent are people naturally Bayesian reasoners? There is currently a raging debate on this issue in psychology (see Stanovich, 2009; Tversky & Kahnman, 1984; vs. J. R. Anderson, 1990; Edwards, 1968; Griffiths et al., 2008; Hahn & Oaksford, 2007), with some taking a middle position (e.g., Juslin, Nilsson, & Winman, 2009; Nickerson, 2004; Reyna, Adam, Poirier, LeCroy, & Brainerd, 2005). Greater attention to alternative framework can help position educational psychologists to make contributions to these issues.

In summary, attention to dialogue theory, Bayesian analysis, and other contemporary frameworks can help move the study of argumentation in education forward. It is time to begin a dialogue among researchers about these dialogic and epistemic concepts.

ACKNOWLEDGMENTS

I thank Ivan V. Ivanov for reviewing previous versions of the article.

REFERENCES

Abrami, P. C., Bernard, R. M., Borokhovski, A. W., Surkes, M. A., Tamim, R., & Zhang, D. (2008). Instructional interventions affecting critical thinking skills and dispositions: A stage 1 meta-analysis. *Review of Educational Research, 78,* 1102–1134.

Alexopoulou, E., & Driver, R. (1996). Small-group discussion in physics: Peer interaction modes in pairs and fours. *Journal of Research in Science Teaching, 33,* 1099–1114.

Anderson, J. R. (1990). *The adaptive character of thought.* Hillsdale, NJ: Erlbaum.

Anderson, J. R., & Lebiere, C. (Eds.). (1998). *The atomic components of thought*. Mahwah, NJ: Erlbaum.

Anderson, R. C., Nguyen-Jahiel, K., McNurlen, B., Archodidou, A., Kim, S., Reznitskaya, A., . . . Gilbert, L. (2001). The snowball phenomenon: Spread of ways of talking and ways of thinking across groups of children. *Cognition and Instruction, 19,* 1–46.

Andriessen, J. (2006). Arguing to learn. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 443–460). New York, NY: Cambridge University Press.

Andriessen, J., Baker, M., & Suthers, D. (Eds.). (2003). *Arguing to learn: Confronting cognitions in computer-supported collaborative learning environments.* Boston, MA: Kluwer.

Andriessen, J., & Coirier, P. (Eds.). (1999). *Studies in writing: Vol. 5. Foundations of argumentative text processing.* (G. Rijlaarsdam & E. Esperet, Series Eds.). Amsterdam, The Netherlands: Amsterdam University Press.

Angell, R. B. (1964). *Reasoning and logic.* New York, NY: Appleton-Century-Crofts.

Asterhan, C. S. C., & Schwarz, B. (2007). The effects of monological and dialogical argumentation on concept learning in evolutionary theory. *Journal of Educational Psychology, 99,* 626–639.

Atkinson, K., Bench-Capon, T., & McBurney, P. (2006). Computational representation of practical argument. *Synthese, 152,* 157–206.

Aulls, M. W. (1998). Contributions of classroom discourse to what content students learn during curriculum enactment. *Journal of Educational Psychology, 90,* 56–89.

Baker, M. (2009). Argumentative interactions and the social construction of knowledge. In N. M. Mirza & A.-N. Perret-Clermont (Eds.), *Argumentation and education: Theoretical foundations and practices* (pp. 127–144). New York, NY: Springer.

Bentham, J. D. (1988). *An introduction to the principles of morals and legislation*. Buffalo, NY: Prometheus Books. (Original work published 1789)

Bell, P., & Linn, M. C. (2000). Scientific arguments as learning artifacts: Designing for learning from the web with KIE. *International Journal of Science Education, 22,* 797–817.

Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Hillsdale, NJ: Erlbaum.

Blair, J. A. (1999). Review of (Walton, 1996). *Argumentation, 13,* 338–343.

Boushey, H., Brocht, C., Gundersen, B., & Bernstein, J. (2001). *Hardship in America: The real story of working families.* Washington, DC: Economic Policy Institute.

Bradley, R. (2003, July). *Probabilism and mental kinematics*. Paper presented at the Remembering Jeffrey Workshop on Bayesian Epistemology, 26th International Wittgenstein Symposium, Kirchberg, Austria.

Brookfield, S. D., & Preskill, S. (1999). *Discussion as a way of teaching.* San Francisco, CA: Jossey-Bass.

Brown, R. A. J., & Renshaw, O. D. (2000). Collective argumentation: A sociocultural approach to reframing classroom teaching and learning. In H. Cowie & G. van der Aalsvoort (Eds.), *Social interaction in learning and instruction: The meaning of discourse for the construction of knowledge* (pp. 52–66). New York, NY: Elsevier Science.

Carbogim, D. V., Robertson, D., & Lee, J. (2000). Argument-based applications to knowledge engineering. *The Knowledge Engineering Review, 15,* 119–149.

Chambliss, M. J., & Murphy, P. K. (2002). Fourth and fifth graders representing the argument structure in written texts. *Discourse Processes, 34,* 91–115.

Chin, C., & Osborne, J. (2010). Supporting argumentation through students' questions: Case studies in science classrooms. *The Journal of the Learning Sciences, 19,* 230–284.

Chinn, C. A. (2006). Learning to argue. In A. M. O'Donnell, C. E. Hmelo-Silver, & G. Erkens, (Eds.), *Collaborative learning, reasoning, and technology* (pp. 355–383). Mahwah, NJ: Erlbaum.

Clark, D. B., & Sampson, V. (2008). Assessing dialogic argumentation in online environments to relate structure, grounds, and conceptual quality. *Journal of Research in Science Teaching, 45,* 293–321.

Copi, I. M., & Cohen, C. (1998). *Introduction to logic* (10th ed.). Upper Saddle River, NJ: Prentice Hall.

Damasio, A. R. (1994). *Descartes' error: Emotion, reason, and the human brain,* New York, NY: Putnam.

De La Paz, S. (2005). Effects of historical reasoning instruction and writing strategy mastery in culturally and academically diverse middle school classrooms. *Journal of Educational Psychology, 97,* 139–156.

Dong, T., Anderson, R. C., Kim, I.-H., & Li, Y. (2008). Collaborative reasoning in China and Korea. *Reading Research Quarterly, 43,* 400–424.

Duschl, R. A. (2008). Quality argumentation and epistemic criteria. In S. Erduran & M. P. Jimenez-Aleixandre (Eds.), *Argumentation in science education: Perspectives from classroom-based research* (pp. 159–175). Dordrecht, The Netherlands: Springer.

Edwards, W. (1968). Conservatism and human information processing. In B. Kleinmuntz (Ed.), *Formal representation of human judgment* (pp. 17–52). New York: Wiley.

Ehninger, D. (1974). *Influence, belief, and argument: An introduction of responsible persuasion.* Glenview, IL: Scott Foreman.

Ehninger, D., & Brockriede, W. (1961). *Decision by debate.* New York, NY: Dodd, Mead.

Erduran, S. (2008). Methodological foundations in the study of argumentation in the science classroom. In S. Erduran & M. P. Jimenez-Aleixandre (Eds.), *Argumentation in science education: Perspectives from classroom-based research* (pp. 47–69). Dordrecht, The Netherlands: Springer.

Erduran, S., & Jimenez-Aleixandre, M. P. (Eds.). (2008). *Argumentation in science education: Perspectives from classroom-based research.* Dordrecht, The Netherlands: Springer.

Erduran, S., Simon, S., & Osborne, J. (2004). TAPping into argumentation: Developments in the application of Toulmin's argument pattern for studying science discourse. *Science Education, 88,* 915–933.

Ferretti, R. P., Lewis, W. E., & Andrews-Wekerly, S. (2009). Do goals affect the structure of students' argumentative writing strategies? *Journal of Educational Psychology, 101,* 577–589.

Fulkerson, R. (1996a). *Teaching the argument in writing*. Urbana, IL: National Council of Teachers of English.

Fulkerson, R. (1996b). The Toulmin model of argument and the teaching of composition. In B. Emmel, P. Resch, & D. Tenney (Eds.), *Argument revisited, argument redefined: Negotiating meaning in the composition classroom* (pp. 45–72). Thousand Oaks, CA: Sage.

Garcia-Mila, M., & Anderson, C. (2008). Cognitive foundations of learning argumentation**.** In S. Southerland (Series Ed.), S. Erduran & M. P. Jiménez-Aleixandre (Volume Eds.), *Argumentation in science education*: *Perspectives from classroom-based research* (pp. 31–48). Dordrecht, The Netherlands: Springer.

Gilbert, M. (1997). *Coalescent argument*. Mahwah, NJ: Erlbaum.

Glymour, C. (2001). *The mind's arrows.* Cambridge, MA: MIT Press.

Golanics, J. D., & Nussbaum, E. M. (2008). Enhancing collaborative online argumentation through question elaboration and goal instructions. *Journal of Computer Assisted Learning, 24,* 167–180.

Graesser, A. C., Baggett, W., & Williams, K. (1996). Question-driven explanatory reasoning. *Applied Cognitive Psychology, 10,* S17–S31.

Greeno, J. G., & van de Sande, C. (2007). Perspectival understanding of conceptions and conceptual growth in interaction. *Educational Psychologist, 42,* 9–23.

Grennan, W. (1997). *Informal logic: Issues and techniques.* Buffalo, NY: McGill-Queen's University Press.

Griffiths, T. L., Kemp, C., & Tenenbaum, J. B. (2008). Bayesian models of cognition. In R. Sun (Ed.), *The Cambridge handbook of computational psychology* (pp. 59–100). New York, NY: Cambridge University Press.

Haake, J. M., & Pfister, H.-R. (2007). Flexible scripting in net-based learning groups. In F. Fischer, I. Kollar, H. Mandl, & J. M. Haake (Eds.), *Scripting computer-supported collaborative learning: Cognitive, computational and educational perspectives* (pp. 155–175). New York, NY: Springer.

Hahn, U., & Oaksford, M. (2006). A Bayesian approach to informal argument fallacies. *Synthese, 152,* 207–236.

Hahn, U., & Oaksford, M. (2007). The rationality of informal argumentation: A Bayesian approach to reasoning fallacies. *Psychological Review, 114,* 704–732.

Hairston, M. C. (1981). *Successful writing: A rhetoric for advanced composition.* New York, NY: Norton.

Halpern, D. F. (1997). *Teaching critical thinking across the curriculum: A brief edition of thought and knowledge.* Mahwah, NJ: Erlbaum.

Hastings, A. C. (1963). *A reformulation of the modes of reasoning in argumentation.* Unpublished doctoral dissertation, Northwestern University, Evanston, Illinois.

Hochs, S. J. (1984). Availability and inference in predictive judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 649–662.

Howson, C., & Urbach, P. (1989). *Scientific reasoning: The Bayesian approach*. La Salle, IL: Open Court.

Hunt, E., & Minstrell, J. (1994). A cognitive approach to the teaching of physics. In K. McGilly (Ed.), *Classroom lessons: Integrating cognitive theory and classroom practice* (pp. 51–73). Cambridge, MA: MIT Press.

Juslin, P., Nilsson, H., & Winman, A. (2009). Probability theory, not the very guide of life. *Psychological Review, 116,* 856–874.

Kant, I. (1996). *Critique of practical reason* (T. K. Abbott, Trans.). Amherst, NY: Prometheus Books. (Original work published 1788)

Kelly, G. J., Drunker, S., & Chen, C. (1998). Students' reasoning about electricity: combining performance assessments with argumentation analysis. *International Journal of Science Education, 20,* 849–871.

Kelly, G. J., Regev, J., & Prothero, W. (2008). Analysis of lines of reasoning in written argumentation. In S. Southerland (Series Ed.), S. Erduran & M. P. Jiménez-Aleixandre (Volume Eds.), *Argumentation in science education*: *Perspectives from classroom-based research* (pp. 137–157). Dordrecht, The Netherlands: Springer.

Kemp, C., & Tenenbaum, J. B. (2009). Structured statistical models of inductive reasoning. *Psychological Review, 116,* 20–58.

Kenyon, L., & Reiser, B. J. (2006, April). *A functional approach to nature of science: Using epistemological understandings to construct and evaluate explanations.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Kienpointner, M. (1992). How to classify arguments. In F. H. van Eemeren, R. Grootendorst, J. A. Blair, & C. A. Willard (Eds.), *Argumentation illuminated* (pp. 178–188). Amsterdam, The Netherlands: Amsterdam University Press.

Kirschner, P. A., Buckingham Shum, S. J., & Carr, C. S. (Eds.). (2003). *Visualizing argumentation: Software tools for collaborative and educational sense-making.* New York, NY: Springer.

Kopp, B., & Mandl, H. (2007, July). Fostering argumentation with script and content scheme in videoconferencing. In C. Chinn, G. Erkens, & S. Puntambekar (Eds.), *Mice, minds, and society: The Computer Supported Collaborative Learning (CSCL) Conference 2007, Vol. 8. Part 1* (pp. 385–394). New Brunswick, NJ: International Society of the Learning Sciences.

Kuhn, D. (1991). *The skills of argument.* Cambridge, England: Cambridge University Press.

Kuhn, D. (2005). *Education for thinking*. Cambridge, MA: Harvard University Press.

Kuhn, D., Goh, W., Iordanou, K., & Shaenfield, D. (2008). Arguing on the computer: A microgenetic study of developing argument skills in a computer-supported environment. *Child Development, 79,* 1310–1328. doi:10.1111/j.1467-8624.2008.01190.x

Kuhn, D., & Udell, W. (2003). The development of argument skill. *Child Development, 74,* 1245–1260.

Kuhn, T. S. (1970). *The structure of scientific revolutions*. Chicago, IL: University of Chicago Press.

Lakatos, I. (1978). *The methodology of scientific research programmes: Philosophical papers* (Vol. 1). New York: Cambridge University Press.

Lampert, M. L., Rittenhouse, P., & Crumbaugh, C. (1996). Agreeing to disagree: Developing sociable mathematical discourse. In D. R. Olson & N. Torrance (Eds.), *Handbook of human development in education* (pp. 731–764). Cambridge, MA: Blackwell.

Leitão, S. (2000). The potential of argument in knowledge building. *Human Development, 43,* 332–360.

Limón, M. (2001). On the cognitive conflict as an instructional strategy for conceptual change: A critical appraisal. *Learning and Instruction, 11,* 357–380.

Linn, M. C., Shear, L., Bell, P., & Slotta, J. D. (1999). Organizing principles for science education partnerships: Case studies of students' learning about "rats in space" and "deformed frogs." *Educational Technology Research & Development, 47*(2), 61–84.

Lovett, M. (2001). A collaborative convergence on studying reasoning processes: A case study in statistics. In S. M. Carver & D. Klahr (Eds.), *Cognition and instruction: Twenty-five years of progress* (pp. 347–384). Mahwah, NJ: Erlbaum.

Mason, L., & Scirica, F. (2006). Prediction of students' argumentation skills about controversial topics by epistemological understanding. *Learning & Instruction, 15,* 492–509.

Mayer, R. E. (1999). *The promise of educational psychology: Learning in the content areas* (2nd ed.). Upper Saddle River, NJ: Merrill Prentice Hall.

McNeil, K. L., & Krajcik, J. (2009). Synergy between teacher practices and curricular scaffolds to support students in using domain-specific and domain-general knowledge in writing arguments to explain phenomena. *The Journal of the Learning Sciences, 18,* 416–460.

Means, M. L., & Voss, J. F. (1996). Who reasons well? Two studies of informal reasoning among children of different grade, ability, and knowledge levels. *Cognition and Instruction, 14,* 139–178.

Metz, K. E. (1998). Emergent ideas of chance and probability in primary-grade children. In S. P. Lajoie (Ed.), *Reflections on statistics: Learning, teaching, and assessment in Grades K–12* (pp. 149–174). Mahwah, NJ: Erlbaum.

Mill, J. S. (1979). *Utilitarianism.* Indianapolis, IN: Hackett. (Original work published 1863)

Muller Mirza, N., & Perret-Clermont, A.-N. (Eds.). (2009). *Argumentation and education: Theoretical foundations and practices*. Dordrecht, The Netherlands: Springer.

Nickerson, R. S. (2004). *Cognition and chance: The psychology of probabilistic reasoning*. Mahwah, NJ: Erlbaum.

Nisbett, R. E. (Ed.). (1993). *Rules for reasoning*. Hillsdale, NJ: Erlbaum.

Nussbaum, E. M. (2003). Appropriate appropriation: Functionality of student arguments and support requests during small-group classroom discussions. *Journal of Literacy Research, 34,* 501–544.

Nussbaum, E. M. (2005). The effect of goal instructions and need for cognition on interactive argumentation. *Contemporary Educational Psychology, 30,* 286–313.

Nussbaum, E. M. (2008a). Collaborative discourse, argumentation, and learning: Preface and literature review. *Contemporary Educational Psychology, 33,* 345–359.

Nussbaum, E. M. (2008b). Using argumentation vee diagrams (AVDs) for promoting argument/counterargument integration in reflective writing. *Journal of Educational Psychology, 100,* 549–565.

Nussbaum, E. M., & Broughton, S. H. (2008, August). *Emotion and attitude change in middle-schoolers' arguments on current events.* Paper presented at the 6th Annual Conference on Conceptual Change, European Association for Research in Learning and Instruction, Turku, Finland.

Nussbaum, E. M., & Edwards, O. V. (in press). Critical questions and argument stratagems: A framework for enhancing and analyzing students' reasoning practices. *The Journal of the Learning Sciences*.

Nussbaum, E. M., & Kardash, C. M. (2005). The effects of goal instructions and text on the generation of counterarguments during writing. *Journal of Educational Psychology, 97,* 157–169.

Nussbaum, E. M., & Schraw, G. (2007). Promoting argument-counterargument integration in students' writing. *The Journal of Experimental Education, 76,* 59–92.

Nussbaum, E. M., Winsor, D. L., Aqui, Y. M., & Poliquin, A. M. (2007). Putting the pieces together: Online argumentation vee diagrams enhance thinking during discussions. *International Journal of Computer-Supported Collaborative Learning, 2,* 479–500.

O'Keefe, D. J. (1982). The concept of argument and arguing. In J. R. Cox & C. A. Willard (Eds.), *Advances in argumentation theory and research* (pp. 3–23). Carbondale, IL: Southern Illinois University Press.

Palincsar, A. M, (1998). Social constructivist perspectives on teaching and learning. *Annual Review of Psychology, 49,* 345–375.

Parsons, S. (2001). *Qualitative methods for reasoning under uncertainty.* Cambridge, MA: MIT Press.

Pearl, J. (2000). *Causality: Models, reasoning, and inference.* New York, NY: Cambridge University Press.

Perelman, C., & Olbrechts-Tyteca, L. (1958/1969). *The new rhetoric: A treatise on argumentation* (J. Wilkinson & P. Weaver, Trans.). Notre Dame, IN: University of Notre Dame Press.

Perkins, D. N. (1985). Postprimary education has little impact on informal reasoning. *Journal of Educational Psychology, 77,* 562–571.

Pollock, J. L. (1987). Defeasible reasoning. *Cognitive Science, 11,* 481–518.

Press, S. J. (2003). *Subjective and objective Bayesian statistics: Principles, models, and applications* (2nd ed.). Hoboken, NJ : Wiley-Interscience.

Pyle, C. H. (2009). *Getting away with torture: Secret government, war crimes, and the rule of law.* Washington, DC: Potomac Books.

Ranney, M., & Schank, P. (1998). Toward an integration of the social and the scientific: Observing, modeling, and promoting the explanatory coherence of reasoning. In S. Read & L. Miller (Eds.), *Connectionist models of social reasoning and social behavior* (pp. 245–274). Mahwah, NJ: Erlbaum.

Reed, C., & Norman, T. J. (2004). A roadmap of research argument and computation. In C. Reed & T J. Norman (Eds.), *Argumentation machines: New frontiers in argument and computation* (pp. 1–13). Boston, MA: Kluwer

Reiser, B., Tabak, I., Sandoval, W., Smith, B. K., Steinmuller, F., & Leone, A. J. (2001). BGuILE: Strategic and conceptual scaffolds for scientific inquiry in biology classrooms. In S. M. Carver & D. Klahr (Eds.), *Cognition and instruction: Twenty-five years of progress* (pp. 263–306). Mahwah, NJ: Erlbaum.

Rescher, N. (1976). *Plausible reasoning: An introduction to the theory and practice of plausibilistic inference*. Assen, The Netherlands: Van Gorcum.

Resnick, L., Salmon, M., Zeitz, C. N., & Wathen, S. H. (1993). Reasoning in conversation. Special issue in discourse and shared reasoning. *Cognition and Instruction, 11,* 347–364.

Reyna, V. F., Adam, M. B., Poirier, K. M., LeCroy, C W., & Brainerd, C. J. (2005). Risky decision making in childhood and adolescence: A fuzzy-trace theory approach. In J. E. Jacobs & P. A. Klaczynski (Eds.), *The development of judgment and decision making in children and adolescents* (pp. 77–106). Mahwah, NJ: Erlbaum.

Reynolds, G. A., & Perin, D. (2009). Comparison of text structure and self-regulated writing strategies for composing from sources by middle school students. *Reading Psychology, 30,* 265–300.

Reznitskaya, A., & Anderson, R. C. (2002). The argument schema and learning to reason. In C. C. Block & M. Pressley (Eds.), *Comprehension-instruction*: *Research-based best practices* (pp. 319–334). New York, NY: Guilford.

Reznitskaya, A., Anderson, R. C., & Kuo, L.-J. (2007). Teaching and learning argumentation. *Elementary School Journal, 107,* 449–472.

Reznitskaya, A., Anderson, R. C., McNurlen, B., Ngyuen-Jahiel, K., Archodidou, A., & Kim, S. (2001). Influence of oral discussion on written argument. *Discourse Processes, 32,* 155–175.

Rottenberg, A. T. (1985). *Elements of argument: A text and reader*. New York, NY: St. Martin Press.

Rowe, G., Macagno, F., Reed, C., & Walton, D. (2006). *Araucaria* as a tool for diagramming arguments in teaching and studying philosophy. *Teaching Philosophy, 29,* 111–124.

Royall, R.M. (1997) *Statistical evidence: A likelihood paradigm.* London: Chapman & Hall.

Sandoval, W. A., & Millwood, K. A. (2005). The quality of students' use of evidence in written scientific explanations. *Cognition and Instruction, 23,* 23–55.

Savage, L. (1972). *The foundations of statistics.* New York: Dover.

Scheuer, O., Loll, F., Pinkwart, N., & McLaren, B. M. (2010). Computer-supported argumentation: A review of the state of the art. *International Journal of Computer-Supported Collaborative Learning, 5,* 43–102.

Schwarz, B. B., Neuman, Y., & Biezuner, A. (2000). Two wrongs may make a right. If they argue together! *Cognition and Instruction, 18,* 461–494.

Sinatra, G. M., & Chinn, C. (in press). Thinking and reasoning in science: Promoting epistemic conceptual change. In K. Harris & S. Graham (Series Eds.) & K. Harris, C. B. McCormick, G. M. Sinatra, & J. Sweller (Volume Eds.), *APA educational psychology handbook series (Vol. 1). Critical theories and models of learning and development relevant to learning and teaching.* Washington, DC: American Psychological Association.

Sinatra, G. M., Southerland, S. A., McConaughy, F., & Demestes, J. (2003). Intentions and beliefs in students' understanding and acceptance of biological evolution. *Journal of Research in Science Teaching, 40,* 510–528.

Stanovich, K. E. (2009). *What intelligence tests miss: The psychology of rational thought.* New Haven, CT: Yale University Press.

Statman, D. (1997). Introduction to virtue ethics. In D. Statman (Ed.), *Virtue-ethics: A critical reader* (pp. 1–37). Washington, DC: Georgetown University Press.

Suedfeld, P., Tetlock, P. E., & Streufert, S. (1992). Conceptual/integrative complexity. In C. P. Smith (Ed.), *Motivation and personality: Handbook of thematic content analysis* (pp. 393–400). New York, NY: Cambridge University Press.

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

Thagard, P. (2004). Causal inference in legal decision making: Explanatory coherence vs. Bayesian networks. *Applied Artificial Intelligence, 18,* 231–249.

Thagard, P. (2006). *Hot thought: Mechanisms and applications of emotional cognition.* Cambridge, MA: MIT Press.

Toth, E. E., Suthers, D. D., & Lesgold, A. M. (2002). "Mapping to know": The effects of representational guidance and reflective assessment on scientific inquiry. *Science Education, 86,* 264–286.

Toulmin, S. (1958). *The uses of argument*. New York, NY: Cambridge University Press.

Toulmin, S. (2003). *The uses of argument* (2nd ed.). New York, NY: Cambridge University Press.

Toulmin, S., Rieke, R., & Janik, A. (1984). *An introduction to reasoning* (2nd ed.). New York, NY: McMillan.

Tversky, A., & Kahnman, D. (1984). Judgment under uncertainty: Heuristics and biases. *Science, 185,* 1124–1131.

van Eemeren, F. H., & Grootendorst, R. (1992). *Argumentation, communication, and fallacies*. Hillsdale, NJ: Erlbaum.

van Eemeren, F. H., Grootendorst, R., Henkemans, F. S., Blair, J. A., Johnson, R. A., Krabbe, E. C. W., . . . Zarefsky, D. (1996), *Fundamentals of argumentation theory*. Mahwah, NJ: Erlbaum.

van Eemeren, F. H., Grootendorst, R., Jackson, S., & Jacobs, S. (1993). *Reconstructing argumentative discourse*. Tuscaloosa: University of Alabama Press.

Vorobej, M. (2006). *A theory of argument.* New York, NY: Cambridge University Press.

Voss, J. F. (Ed.). (2001). Argumentation in psychology [Special issue]. *Discourse Processes, 32,* 2/3.

Voss, J. F., Tyler, S. W., & Yengo, L. A. (1983). Individual differences in the solving of social science problems. In R. D. Dillon & R. R. Schmeck (Eds.), *Individual differences in cognition* (Vol. 1, pp. 205–235). Cambridge, MA: MIT Press.

Vye, N. J., Goldman, S. R., Voss, J. F., Hmelo, C., Williams, S., & the Cognition and Technology Group at Vanderbilt. (1997). Complex mathematical problem solving by individuals and dyads. *Cognition and Instruction, 15,* 485–484.

Walton, D. (1992). *The place of emotion in argument.* University Park: The Pennsylvania State University Press.

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

Walton, D. N. (1998). *The new dialectic.* University Park: Pennsylvania State University Press.

Walton, D. (2000). The place of dialogue theory in logic, computer science and communication studies. *Synthese, 123,* 327–346.

Walton, D. (2003). *Ethical argumentation*. Lanham, MD: Lexington Books.

Walton, D. (2004). *Abductive reasoning.* Tuscaloosa: University of Alabama Press.

Walton, D. (2007). *Dialogue theory for critical argumentation*. Philadelphia, PA: John Benjamins.

Walton, D. (2010). A dialogue model of belief. *Argument and Computation, 1,* 23–46.

Walton, D., Reed, C., & Macagno, F. (2008). *Argumentation schemes*. New York, NY: Cambridge University Press.

Webb, N. M., Franke, M. L., Ing, M., Chan, A., De, T., Freund, D., & Battey, D. (2008). The role of teacher instructional practices in student collaboration. *Contemporary Educational Psychology, 33,* 360–381.

Wegerif, R., Mercer, N., & Dawes, L. (1999). From social interaction to individual reasoning: An empirical investigation of a possible socio-cultural model of cognitive development. *Learning and Instruction, 9,* 493–516.

Weinberger, A., Stegmann, K., & Fischer, F. (2005). Computer supported collaborative learning in higher education: Scripts for argumentative knowledge construction in distributed groups. In T. Koschmann, D. Suthers, & T. W. Chan (Eds.), *Proceedings of the International Conference on Computer Supported Collaborative Learning CSCL 2005* (pp. 717–726). Mahwah, NJ: Erlbaum.

Wiley, J., & Voss, J. F. (1999). Constructing arguments from multiple sources: Tasks that promote understanding and not just memory for text. *Journal of Educational Psychology, 91,* 301–311.

Williamson, J. (2005). *Bayesian nets and causality*. New York, NY: Oxford University Press.

Windschitl, M., & Thompson, J. (2006). Transcending simple forms of school science investigation: The impact of preservice instruction on teachers' understandings of model-based inquiry. *American Education Research Journal, 43,* 783–835.

Yeh, S. S. (1998). Empowering education: Teaching argumentative writing to cultural minority middle school students. *Research in the Teaching of English, 33,* 49–83.

APPENDIX

BAYES' THEOREM EXPLAINED: THE MARY EXAMPLE

This appendix provides a technical explanation of Bayes' theorem and how it can be applied to argumentation using an extended, hypothetical example regarding a woman named Mary. Mary and her husband have been trying to conceive a child for the better part of a year but have been unsuccessful. Although Mary thinks it is unlikely that she is pregnant, one day she takes a store-bought pregnancy test and it comes out positive, much to her delight.

One could now make an argument that Mary is likely pregnant, because the test result was positive. If one were to use the Toulmin model, this argument could be structured as:

(1) Mary is probably pregnant (claim), because
(2) The pregnancy test was positive (grounds), and
(3) A positive test usually indicates one is pregnant (warrant),
(4) Based on studies by the manufacturer of the reliability of the pregnancy test (backing),
(5) Unless the test was misread or otherwise yielded a false positive. i.e., comes out positive when Mary is not pregnant (rebuttal).
(6) Therefore, Mary is probably pregnant but we cannot be absolutely sure (qualifier).

A Bayesian approach would attach specific probabilities to the various argument parts. It would also recognize that Mary had an initial belief, before taking the test, regarding the probability that she was pregnant. Following is the argument cast in Bayesian terms:

(1) Although we initially thought that Mary is not pregnant (our subjective *prior probability* that she is pregnant was only 40%),
(2) The pregnancy test was positive (*evidence*), and
(3) If somebody is pregnant, the test comes out positive 80% of the time (*likelihood*),
(4) Based on studies conducted by the manufacturer (*how likelihood was set*),
(5) But when someone is not pregnant, the test comes out positive 5% of the time (*false positive rate*),
(6) So we conclude, with 91% certainty, that Mary is probably pregnant (*posterior probability*).

This example attaches specific probabilities to each premise as well as to the conclusion. The probability of the conclusion is called the *posterior probability* because it is the probability after the evidence is considered. The *prior* probability was 40%, but now—given the evidence—the probabilistic strength in the belief that Mary is pregnant should now rise to 91%. However, where does the 91% cited in the conclusion come from? It is derived from Bayes' theorem, which is a formula for deriving the posterior probability given the following three parameters:

- Prior probability,
- Likelihood, and
- False positive rate.

In the previous argument, Bayes' theorem is an additional warrant justifying the probability estimate in the conclusion. The specific calculations, as well as Bayes' theorem, are presented in Table A1. Table A2 presents the same calculations in the form of a contingency table for 100 hypothetical women (starting with an incomplete table with just information on the three parameters). Table A3 presents the same analysis in terms of odds, where is it shown that the odds that Mary is pregnant are a function of weighing the likelihood against the false positive rate. (This is analogous to weighing arguments against counterarguments.) In all three cases, although the probability (or odds) was initially against Mary being pregnant, with the introduction of new evidence (i.e., the results of the pregnancy test), the probability (or odds) are now highly favorable.

TABLE A1
Bayes' Theorem Applied to the Mary Argument

1. Definitions: Let X be the event that Mary is pregnant, P(X) the prior probability that she is pregnant (strength of our belief before examining evidence), E the evidence (e.g., that the test comes out positive), P(E | X) the likelihood (that the test is positive given that Mary is pregnant), P(E | ∼ X) the false positive rate (the probability that the test is positive given than Mary is not pregnant), and P(X | E) the posterior probability (what the probability should be, given the evidence).
2. Bayes' Theorem[a]:
   a. $\dfrac{Likelihood * prior}{Likelihood * prior + Falsepositiverate * (1 - prior)} = Posterior\ probability$
   b. Written symbolically: $\dfrac{P(E|X) * P(X)}{P(E|X)P(X) + P(E|\sim X)P(\sim X)} = P(X|E)$
3. Parameters:
   c. Prior probability belief: 40%.
   d. Likelihood: 80%
   e. False Positive Rate: 5%.
4. Calculation:
   $$\frac{80\% * 40\%}{80\% * 40\% + 5\% * 60\%} = \frac{32\%}{32\% + 3\%} = 91\%$$
5. Conclusion:
   The posterior probability is 91%. Mary should increase her estimate of the probability that she is pregnant from 40% to 91%, given the new evidence.

[a] Bayes' theorem is derived from the definition of conditional probabilities: P(A | B) = P(A & B)/P(B). Also, P(B | A) = P(A&B)/P(A), so P(A&B) = P(B | A)P(A). Substituting for P(A&B) in the first equation yields P(A | B) = P(B | A)P(A)/P(B), which is Bayes' theorem.

TABLE A2
Representation of Mary Argument Probabilities With Contingency Tables ($N = 100$)

Representation of Mary argument probabilities with contingency table partially filled in with Bayesian parameters

	Pregnant	Not Pregnant	Total
+Test	32[a]	3[b]	
–Test			
Total	40[c]	60	100

[a]The conditional probability that there is a positive test result, given that someone is pregnant, is $\frac{32}{40} = 80\%$. [b]The conditional probability that there is a positive test result, given that someone is pregnant, is $\frac{3}{60} = 5\%$. [c]Prior probability is 40%.

Representation of Mary argument probabilities with contingency table completely filled in

	Pregnant	Not Pregnant	Total
+Test	32	3	35
–Test	8	57	65
Total	40	60	100

Note. Probability that someone is pregnant, given a + test result, is $\frac{32}{35} = 91.4\%$.

TABLE A3
Odds Version of Bayes' Theorem

1. Bayes' theorem can be recast in terms of odds. Although mathematically equivalent to the formula in Table A1, the odds version is more intuitively understandable.
2. The odds of an event is defined as:

   $Odds(X) = \frac{P(X)}{1-P(X)}$

   The prior odds that Mary is pregnant are:

   $\frac{40\%}{(1-40\%)} = 0.66$

   The posterior odds that Mary is pregnant are:

   $\frac{91\%}{(1-91\%)} \approx 10 \text{ to } 1$
3. The odds version of Bayes' theorem is:

   $Prior\ odds = \left(\frac{Likelihood}{False\ Positive\ Rate}\right) = Posterior\ odds$

   The theorem involves weighing the likelihood against the false positive rate, which is analogous to weighing arguments against counterarguments (the counterargument being that the reading could be a false positive).
4. Applied to the Mary argument, the formula is:

   $\frac{80\%}{5\%} * 0.66 = 10.66$ same as the previous calculation.
5. Conclusion:

   Although the odds were initially against Mary being pregnant (i.e., odds were less than one), with the new evidence, the odds that she is pregnant are highly favorable (about 10 to 1).