May 17th, 9:00 AM - May 19th, 5:00 PM

Sampling scholarly arguments: a test of a theory of good inference

David Hitchcock
McMaster University

May 17th, 9:00 AM - May 19th, 5:00 PM

Sampling scholarly arguments: a test of a theory of good inference

David Hitchcock
McMaster University

Title: Sampling scholarly arguments: a test of a theory of good inference
Author: David Hitchcock
Response to this paper by: Jonathon Adler

For some years, I have been developing a general theory of inference appraisal (Hitchcock 1985, 1992, 1994, 1998). According to a current version of this theory, a conclusion follows (conclusively, provisionally, etc.) from premisses adduced in its support if and only if some covering generalization of the argument is non-trivially acceptable (always, provisionally, etc.). By a *covering generalization* is meant a generalization of the conditional sentence whose antecedent is the conjunction of the premisses and whose consequent is the conclusion. Such a generalization is *acceptable* if and only if it deserves to be accepted by the appraiser; criteria of acceptability may vary between appraisers and between different types of sentences (empirical, normative, evaluative, etc.). The acceptability of a generalized conditional is *non-trivial* if it is based neither on the (universal, provisional, etc.) unacceptability of its antecedent nor on the (universal, provisional, etc.) acceptability of its consequent.

This admittedly schematic theory gives us what we want out of a good inference: transmission of acceptability from premisses to conclusion. It covers conclusive as well as non-conclusive inferences. It captures the generality of inference claims which is implicit in the human practice of refutation by logical analogy. It covers speech acts for which truth seems an inapplicable criterion of acceptability–such as recommendations, evaluations and requests. It provides a touchstone for testing proposed reasoning schemes or argument schemes expressed at a lower level of generality. And it fits our practices of reasoning and argument much better than the widely accepted view that good inference is a matter of formal deductive validity or conformity to canons of a formal inductive logic.

For example, looking at the grey sky and blowing tree branches outside my window, I may infer that today's forecast of rain is likely to be correct. If I were to verbalize this inference, I would do so somewhat as follows: "It probably will rain today: the sky is a leaden grey, and the tree branches outside my window are blowing furiously in the wind." I would not include an extra premiss which would make my reasoning formally valid, such as: "Whenever the sky is a leaden grey, and the tree branches outside my window are blowing furiously in the wind, it is going to rain." This covering generalization is non-trivially true for the most part, as we know by experience and elementary meteorology. Its acceptability means that, according to the theory sketched above, the conclusion of my inference follows from the premisses just as they are stated, without supplementation.

One way to test any such theory is to apply it to actual reasonings. For this purpose, it is desirable to have a sample which is likely to be representative of the variety of ways in which human beings draw conclusions. I used random methods to extract a sample of 50 inferences from several hundred thousand English-language books catalogued in the library of a research-intensive university. In this paper I describe the method of sampling, the criteria for identifying inferences, and the results of applying the above theory to the 50 inferences.

1. Method

We selected at random a starting-point within the first 50 lines of each of the first 500 pages of the English-language books catalogued in the libraries of McMaster University. The computerized database of the university's library holdings contained at the start of the project

1,204,802 entries numbered sequentially. A random number generator was used to generate numbers between 1 and 1,204,802 which were used to identify the work from which a selection was to be made. It was then used to generate a random number between 1 and 500, for the starting page, and then a random number between 1 and 50, for the starting line. (If a page was printed in two columns, we counted two lines in a column as one; for example, if the random number generator produced the number 12 for a double-column page, we started at the 23rd line of the first column.) If at any of these three stages of selection an unacceptable or non-existent item was selected, then the search was stopped at this point and the next number pursued. The search stopped at the first stage if the work selected was not an English-language publication or was a periodical. It stopped at the second stage if the page number selected was greater than the number of the last page in the work. It stopped at the third stage if the line number selected was greater than the number of the last line on the page.

Two of us (Craig Faucette, my research assistant, and I) independently read forward from the beginning of the line so identified until we found the first complete inference or argument. If the starting-point happened to be in the middle of an inference or argument, then we went forward to the next inference or argument; in other words, we looked for the first inference or argument which started at or after the designated starting-point. If we got to the end of the book without finding an inference or argument, then we went on to the next book in the numbered series, so that there was no bias against inferences or arguments at the beginning of a book. Before beginning our work, we discussed the concepts of inference and argument to be used, and practised identifying them until we got reasonable agreement. (Our practice text was Jürgen Habermas' *Theory of Communicative Action*, which Faucette was reading for a master's thesis he was writing.)

We used the following concepts of argument and inference, and the following criteria for detecting and standardizing them. A verbal inference is a discourse in which someone draws for himself or herself a conclusion on the basis of one or more premisses; the person drawing the conclusion may be either the author of the text or a character in the text whose discourse is being reported, by quotation or paraphrase. Such a person could be said to state what he or she thinks, then explain why he or she thinks this; the supporting reason(s) could of course precede the conclusion in the text. An argument is a discourse in which someone invites hearer(s) or reader(s) to make an inference; the author of the argument may be the author of the text in which the argument occurs, but reported, quoted or fictionally imputed arguments also counted, as long as they were articulated in full detail. We extracted the inference or argument in the following standard form:

Premiss(es):
Conclusion:

We wrote the conclusion and each premiss exactly as they occur in the text, with the following modifications. We omitted inferential indicators such as *since* and *therefore*, and also modal qualifiers like *must* and *seems* which indicate the claimed strength of the inference from premiss(es) to conclusion. We supplied in square brackets contextual information needed to clarify the reference of pronouns, noun phrases and so. We put material interpolated to fill out an elliptical text in angle brackets. Otherwise, we made no additions to or subtractions from what is written in the text.

We developed some criteria for deciding on difficult cases. These criteria indicate how we applied the concepts defined above. Their articulation here is a contribution to the art of identifying arguments, which is a more difficult skill than one might imagine. (The examples

used to illustrate the criteria come from classical texts which I used with another research assistant, Darcy Otto, as a basis for getting inter-rater reliability on another task of sampling arguments.)

The mere fact that two nearby clauses in a text could be used as premiss and conclusion is not enough to show that they are being so used. There must be explicit indications in the text that one is being used as a basis for the other, or at least an absence of an explicit counter-indication. The clearest explicit indication is the occurrence of an inferential particle or phrase joining the two clauses, either a premiss indicator like *since* or *given that* preceding the apparent premiss or a conclusion indicator like *therefore* or *it follows that* preceding the apparent conclusion. A clear tripartite indication is that (a) one clause immediately follows the other with no joining inferential particle, (b) the context makes clear that the putative conclusion is not already established or accepted before the putative premiss is mentioned, and (c) it is at least plausible that one could infer the putative conclusion from the other clause; most typically in such cases the conclusion will be stated first, with the premiss following as a supporting reason. (A putative conclusion may be not already established or accepted even though some argument has been given for it; this happens when an author gives a series of independent arguments for the same conclusion, as Thomas Aquinas does for example in his *Summa Theologica* or Aristotle does in various works.)

One explicit counter-indication of inference or argument is the occurrence of a contrastive particle like *but* or *although* joining the two clauses. An author who is contrasting two pieces of information is not using one of them as a basis for drawing the other as a conclusion; see, for example, Homer's *Iliad* 5.150-151, where the first clause could otherwise be taken as a conclusion supported by the second. (Note however that a contrastive particle can occur between two premisses of an argument or inference.)

Another explicit counter-indication of inference or argument is that the putative conclusion has already been accepted by its supposed recipient before the putative premiss is stated. For example, at Plato's *Theaetetus* 177b1, Socrates' statement "I know" could be used as a premiss to get his interlocutor Theodorus to accept his description of how wicked people will react to his description of their fate (it could, for example, be read as an appeal by Socrates to his personal experience of how such people have reacted in the past to his having said such things), but Theodorus has just accepted Socrates' point in the immediately preceding line (177a9), so there is no argument here.

Another counter-indication is that the supposed premiss is the antecedent of a conditional whose consequent is the supposed conclusion. A conditional sentence of the form *If p then q*, whether indicative or contrary-to-fact, is almost never in itself an argument or inference. (The rare exceptions are texts where the context makes clear that the condition in the *if* clause is being tacitly asserted.) Thus, Plato's *Theaetetus* 193e1-4, which has the force of a conditional, is not an argument, just an indicative conditional statement. Similarly, Plato's *Phaedo* 60c1-5, which is a contrary-to-fact conditional, is not an argument. Conditional statements can of course be parts of arguments or inferences, either as a premiss or as a conclusion; they are just not by themselves arguments or inferences.

In general, the fact that the text could be supplemented in such a way as to make an inference or argument out of its components does not make that text an inference or argument. Thus, at Plato's *Sophist* 224a the visitor from Elea mentions all kinds of music, painting and shows which are brought from one city to another for sale. "The one who transports and sells [such things–DH] is no less rightly called *trader* than the seller of food and drink." (*Sophist*

224a6-7) One could support this claim by pointing out that both types bring a product from other cities and sell it. But the stranger does not offer any such support, perhaps taking it as so obvious that he does not need to supply it. Thus there is no argument (or inference) in this text.

A complication in relying on the presence of an inferential particle as a criterion for the presence of inference or argument is that in most languages such particles can function also to indicate a causal explanation. In English, the particle *because* is especially ambiguous in this respect. Suppose that we have a discourse like a complex *because* sentence in English, where one clause (introduced by *because* or its analogue) is subordinate and the other clause is the main clause. In an inference the main clause is drawn as a conclusion from the subordinate *because* clause; the author of the inference thinks that the subordinate clause provides enough evidence to justify accepting the main clause. (In an argument, the arguer expects the intended audience to think this.) In a causal explanation, on the other hand, the author thinks that the subordinate clause gives information about how it came about that the information in the main clause is true. It follows that the main clause of a causal explanation must be a factual claim in the past or present tense; if the main clause expresses a contrary-to-fact circumstance or a wish, or is a directive of some sort, or predicts some future state of affairs, the discourse is an inference or argument, not a causal explanation. See for example *Theaetetus* 177b7-8, where the main clause is a hortatory subjunctive, so that the particle *epeidê* must be inferential rather than causal. Similarly, in Homer's *Iliad*, 9.376-377 the particle *gar* is inferential rather than causal, because the main clause is an imperative.

Where the main clause in a complex sentence with an ambiguous particle is a present or past indicative, the decision as to whether the discourse is a causal explanation or an inference/argument rests on a number of factors, none of which is in itself decisive. The first factor is whether in the context the intended recipient of the discourse (who may also be its author, if the person is simply articulating their own thinking) already accepts the truth of the main clause. If so, it is probably a causal explanation; if not, it is probably an inference. The second factor is whether the information in the subordinate clause works better as a causal explanation of the information in the main clause or as supporting evidence that this information is correct. A good way to test for this factor is to take the main clause by itself and ask oneself in turn:

1. What sorts of events or actions could bring about the state of affairs described by this clause?
2. What sort of evidence could show that this clause is correct?

If the subordinate clause has information of type (1), the discourse is probably a causal explanation. If it has information of type (2), the discourse is probably an inference or argument. The third factor is whether the main clause or the subordinate clause is easier to take for granted. If the subordinate clause is easier to take for granted, then the discourse is probably an inference or argument. If the main clause is easier to take for granted, then the discourse is probably a causal explanation. Thus, for example, in Homer's *Iliad* there is a complex sentence at 9.398-405, with the subordinate conjunction *gar*, "for". In the main clause Achilles, the speaker, tells Odysseus that in his native Phthia he (Achilles) often wanted to get married and enjoy the possessions of his father, ruler of Phthia. In the subordinate clause Achilles says that to him the great wealth of Troy is not worth as much as life. The context is the Achaean siege of Troy, in which Achilles risks his life but stands to gain great riches from plundering the city. Achilles' statement of what he wanted in the past is based on his memory, and is likely to be accepted by Odysseus on Achilles' say-so. So on the first criterion the discourse is probably a causal

explanation. Furthermore, Achilles' value system works well as a causal explanation of what he wanted, but is not very good evidence that he had this desire; the sort of evidence that would establish his having this desire is that he talked to others about marrying and settling down, that he showed an active interest in eligible prospective marriage partners, and so on. So on this factor too the discourse is probably a causal explanation rather than an argument. The third criterion is difficult to apply, and if anything points weakly in the opposite direction: it seems slightly easier to take for granted that Achilles values life more than even the great wealth of Troy than that he often wanted to get married and enjoy his father's possessions. So the preponderance of evidence indicates that this discourse is a causal explanation.

The same points can be made about other types of discourse which in the abstract are ambiguous between an inference or argument and a causal explanation (e.g. an English discourse in which two clauses are joined by *so*) by making appropriate substitutions (e.g. the clause governed by *so* corresponds to the main clause in a *because* sentence).

In cases where there is no inferential particle, the main indication of an inference or argument is that, of two clauses immediately adjacent to one another, one is in the context in need of evidential support and the other gives, or could be plausibly construed by the author of the inference or argument to give, the required evidential support. Thus, for example, in Plato's *Phaedo*, at 60b3-c1, Socrates' exclamation of how strange pleasure seems to be and how surprising its natural relation to its apparent opposite pain involves a puzzling claim, which needs explication and support. It is followed immediately by a pair of contrasted observations– that they don't occur simultaneously in the same person, but that someone who pursues one of them and acquires it is almost always bound to acquire the other one too, as if they were two creatures with one head. These observations can be used as evidence to support a claim that pleasure has a surprising natural relation to pain, the surprise being signalled by the use of a contrast term. Thus it is reasonable to assume that Socrates is in fact using it for just that purpose, to explain why he thinks pleasure and pain have a surprising relationship; the passage is an inference.

We found some inferences and arguments which were part of a chain of reasoning, sometimes a lengthy one. In such cases we included only that part of the chain which we encountered first in reading forward from our starting-point. If the chain of reasoning did not proceed sequentially in the text (for example, because an ultimate conclusion was followed by an ultimate premiss, from which an intermediate conclusion was then derived, as at Aristotle's *Metaphysics* XI.1.1059b24-27), then we picked the first fully completed inference in the chain; if more than one inference ended at the same point in the text, as in the passage just mentioned, then we picked that inference which started first. Thus each text selected includes just one inference from premiss(es) to conclusion. But the description of the context indicates whether the selected text is part of a larger chain of reasoning and, if so, how in broad outline the reasoning goes.

A difficulty in applying this criterion is that it is sometimes difficult to decide how much of the text to include in the premisses. At Plato's *Sophist* 223e5-6, a visitor from Elea responds to Theaetetus' request for an explanation of what he just said with the following inference, nicely signalled in the Greek by the premiss indicator *epei* (*since*): "Perhaps we are ignorant of the one [the kind of trading–DH] concerning the soul, since we understand somewhat the other one [the trading of things for the nourishment and use of the body–DH]." (*Sophist* 223e5-6) To understand this argument, we have to supply the presupposition of Theaetetus' question, "What do you mean by this?" (e4): We [the audience, including Theaetetus] do not understand this. And

the word *this* refers to the visitor's previous statement: "Don't we think then that one kind of trading buys and sells things which the body feeds on and uses, and the other <buys and sells things which> the soul <feeds on and uses>." (e1-3) Thus, a fully informative standardization of the argument, with no alteration of the text, would look as follows:
Premisses: What do you mean by this [that one kind of trading buys and sells things which the body feeds on and uses, and the other <buys and sells things which> the soul <feeds on and uses>–DH]?
We understand somewhat the other one [the trading of things for the nourishment and use of the body–DH].
Conclusion: Perhaps we are ignorant of the one [the kind of trading–DH] concerning the soul.
If we leave out the question, on the ground that it is part of the context of the inference, rather than of the inference itself, then the argument starts at line 5. If we include the question, then the inference starts at line 4. But the question uses an anaphoric reference in the word *this* which point back to the previous three lines; if the lines referred to are counted as part of the argument, then the inference starts at line 1. If the starting-point from which we began to look for arguments and inferences were between lines 1 and 5 inclusive, then the point where we count this inference as beginning would make a difference to whether it was included in our sample. In order to treat such cases uniformly, we decided that lines of text incorporated into an argument or inference by anaphoric reference do not in themselves form part of the inference or argument. Thus, in the above example, lines 1 to 3 are not part of the inference, although they are referred to at line 4. As for line 4, it is implausible to take a question as a basis for drawing an inference; the visitor from Elea has rather interpreted the question as an indication of lack of understanding of his statement. Such implicit components of the text are, on the principles already laid down, not to be included in the inference extracted from it, although they should be mentioned in the description of the context. Thus in this case the inference begins at line 5.

In general, we did not impute inferences and arguments when the putative conclusion was unstated. Allowing inferences and arguments with unstated conclusions opens the door to reading inferences and arguments into texts, because the reader can draw inferences from what is in the text. It also risks building into the sample assumptions about the consequence relation, which would make the whole exercise circular. The conclusion had to be stated, unless the text makes very clear what conclusion is supposed to be drawn from it. As it turned out, the conclusion was explicitly stated in all 50 selected passages.

If the two of us differed on the identification or standardization of an argument or inference, we tried first to resolve our disagreement through discussion. If we did not rapidly reach wholehearted agreement, a third person (Darcy Otto) who had been similarly trained in identifying arguments and inferences independently read the same text from the same starting-point. If he agreed with one of us, I took his judgement as decisive. There were a few occasions when he came up with yet a third suggestion, and the disagreement was resolved through discussion.

All three of us were aware of the theory of consequence whose applicability we were using the sample to test. Awareness of this theory may of course have biased our process of identification.

2. Results

We found 50 inferences and arguments from 28 fields, ranging alphabetically from anthropology to vocational guidance. It was possible to apply the above-mentioned theory of inference

appraisal directly in 49 of the 50 cases, and to reach a judgement of the strength of the link between stated premiss(es) and conclusion. In 12 of the 49 cases, it was necessary to supplement the stated premisses with additional information found in the immediate context or with background information which the intended audience would be able to supply. Apart from this supplementation, application of the theory of inference appraisal almost always involved substantive judgements about the acceptability of a covering generalization. Details of the analysis and evaluation of the 50 passages can be found in the Appendix to this paper.

Table 1. Characteristics of a sample of 50 arguments and inferences

#	Cue	#Pr	S?	C?	V	A?	Type	Field	Str	E?	N?	A/i	D/r
1		3			c	y	ec	biog		0i	s0i	a	d
2		1			m		mepr	eeng				a	d
3		1			i		ibe	trav	n			i	r
4	c/pa	1		y	s		md	math				a	d
5	p/pa	1			i	y	ec	econ				a	d
6	c/pa	1		y	s	y	md	msci				a	d
7	c/v	100			s		qdsd	dst		0	i	a	d
8	p/pa	1			s		ca	pols				a	d
9		5			c		phd	pols	p	w	w	a	d
10	c/pa	1			m	y	mepr	sw				a	d
11	p/pa	2			i		ec	dst		w	i	a	d
12		1			s	y	it	hom				a	d
13		3		y	s		se	fic		0	0i	i	r
14		3			c		mepr	biog	n	i	i	i	r
15	c/v	3		y	i		aa	phy	n	i	i	a	d
16		2			c		pcdm	hist		w	w	i	r
17	p/pa	1	1		i		if	phil	n			a	d
18	p/pu	1			m		if	phil				a	d
19	p/pa	2			m		dco	arth		i	i	a	d
20		6			i		pce	pols		0w	0w	a	d
21	c/v	2	1		p		ipe	psy	s	w	i	a	d
22		2			m		ec	hist		i	i	a	d

23	p/pu	1			i		it	pols				a	d
24	p/pa	2			c		ec	theo		i	i	a	d
25		4			c	y	gi	hist		w	w	a	d
26		5			f	y	eg	pa		0	0	a	d
27		3			i	y	ec	fic		i	i	a	r
28		1			i		it	hist				a	d
29		4			c		cc	biog		w	w	a	d
30	p/pa	3			c		ec	pols		w	w	a	d
31		4			i		it	hist		i	i	a	d
32		5			m		ec	hist		0i	0i	i	r
33	p/ph	4			m		ec, eg	hist		0	0	a	d
34		2			i		ec	hist		i	i	a	d
35	p/pa	4			c	y	mepr	surg		i	i	a	d
36	p/pa	1			i	y	adc	theo				a	d
37		2			m		ec	hist		0	0	a	d
38	c/pa	1			m		dco	ceng				a	d
39	c/pa	6			m		gi	soc		0w	0w	a	d
40	p/pa	1			m	y	mepr	mc	n			a	d
41		3			f		r	pols		0	iw	a	r
42		2			m	y	qdnd	litc		iw	iw	a	d
43		3			m		qdnd	biog		w	w	a	d
44	p/pp	1			s		qdsd	chm				a	d
45		2			m		ce	vocg		w	w	a	d
46	p/ph	3			f		eg	bus	s	0i	0i	a	d
47		1			i	y	eg	phil	s			a	d
48	p/pu	1			m		dco	bib				a	d
49		5			m		ce	anth		0w	0w	a	d
50	c/pa p/v	2		y	m		md	math		i	i	a	d

Legend: # = number of passage. *Cue* = verbal cue to presence of inference or argument: blank = none, p = premiss indicator, c = conclusion indicator, pa = particle (e.g. “hence”), ph = phrase

(e.g. “for example”), pp = preposition (e.g. “as indicated by”) pu = punctuation (colon), v = verb (e.g. “reveals that”). *#pr* = number of premisses. *S?* = number of premisses supported by argument: blank = 0. *C?* = inference drawn from conclusion?: y = yes, blank = no. *V* = validity status: c = valid ceteris paribus, f = formally valid, i = invalid, m = materially valid, p = probabilistically valid, s = semantically valid. *A?* = premisses supplemented for evaluation?: y = yes, blank = no. *Type* = type of argument (i.e. argument schema): aa= argument by analogy, adc = application of definition to a case, ca = conceptual analysis, cc = classification by criteria, ce = cause-to-effect reasoning, dco = direct conclusion from observations, ec = evaluation by criteria, eg = existential generalization, gi = generalization from instances, ibe = inference to the best explanation, if = imputation of fallacy, ipe = inference to a possible explanation, it = interpretation of text(s), md = mathematical deduction, mepr = means-end practical reasoning, pcdm = pros-and-cons decision-making, pce = pros-and-cons evaluation, phd = prediction from historical data, qdnd = qualitative description of non-statistical data, qdsd = qualitative description of statistical data, r = repetition, se = semantic entailment. *Field* = field to which book belongs: anth = anthropology, arth = art history, bib = bibliography, biog = biography, bus = business, ceng = computer engineering, chm = chemistry, dst = descriptive statistics, econ = economics, eeng = environmental engineering, fic = fiction, hist = history, hom = homiletics, litc = literary criticism, math = mathematics, mc = museum curatorship, msci = mechanical sciences, pa = public administration, phil = philosophy, phy = physics, pols = political science, psy = psychology, soc = sociology, surg = surgery, sw = social work, theo = theology, trav = travel memoirs, vocg = vocational guidance. *Str* = strength of link claimed by author: n = necessitation, p = probabilification, s = suggestion, blank = no claim. *E?* = effect on link strength of eliminating a premiss: 0 = no change, i = invalidation, w = weakening, blank = irrelevant (only one premiss); more than one entry reflects different effects for different premisses. *N?* = effect on link strength of denying a premiss: 0 = no change, i = invalidation, s = strengthening, w =weakening, blank = irrelevant (only one premiss); more than one entry reflects different effects for different premisses. *A/i*: a = argument, i = inference. *D/r*: d = direct, r = reported.

Table 1 sets out in summary form the characteristics of the 50 arguments and inferences. Only 3 of the inferences (6 %) were formally valid; of these, one was a case of repetition, and the other two cases of existential generalization. In fact, since there was some rewording of the content in each of these 3 cases, they could as easily have been classified as semantically valid. 7 of the inferences (14%) were semantically valid, in the sense that they had a covering generalization which was true in virtue of the meaning of its constituent terms. 17 (34%) were materially valid, in the sense that they had a substantive covering generalization which was acceptable without exceptions. (My use of the phrase *materially valid* for such a relationship is non-standard; I use it because I have found that the most obvious alternative, *enthymematically valid*, is misinterpreted to mean “valid once one adds an unstated premiss,” which is not the concept I have in mind.) Only one (2%) was probabilistically valid, in the sense of having a covering generalization which was true for the most part. 9 (18%) were valid provisionally or *ceteris paribus*, in the sense that they had a covering generalization which was acceptable but rebuttable by exception-making circumstances whose frequency was undeterminable. The remaining 13 inferences (26%) were invalid; that is, no covering generalization was acceptable, even provisionally or probabilistically.

Table 2. Validity status of a sample of 50 inferences and arguments		
Status	Number	Percentage
formally valid	3	6
semantically valid	7	14
materially valid	17	34
probabilistically valid	1	2
valid ceteris paribus	9	18
invalid	13	26

It should be emphasized that considerable judgement was involved in evaluating the inferences. The judgements are mine alone; others may apply the same theory but reach different verdicts. Readers are invited to do their own appraisal of the inferences in the Appendix.

Table 1 describes each passage on a number of dimensions. The distribution of values on each of these variables is of some relevance to the pedagogy of argument analysis and evaluation, though perhaps more for the sense it gives of the range of argumentative and inferential texts to be found in the books of a university library than for the percentages involved; see the following section for comments on the large margin of error and possible unrepresentative of this sample. I give here the results with respect to each dimension.

Table 3. Verbal cues to the presence of argument/inference in a sample of 50 passages							
None	Premiss indicator (17 [34%])					Conclusion indicator (9 [18%])	
25 (50%)	particle	verb	preposition	phrase	colon	particle	verb
	10 (20%)	1 (2%)	1 (2%)	2 (4%)	3 (6%)	6 (12%)	3 (6%)

Presence of explicit cues to the presence of an argument or inference: The passages divided exactly equally into those with an explicit verbal indication of the presence of an inference or argument, and those without any such indication. Of the 25 with an explicit verbal indication, 17 (34%) contained a premiss indicator; of these 17, 10 (20% of the total) were particles like *since* or *because*, 1 (2%) was the verb *follows from*, 1 (2%) was the prepositional phrase *as indicated by*, 2 (4%) were parenthetical phrases like *for instance*, and 3 (6%) were punctuation marks (a colon, in each case). 9 of the passages (18% of the total) contained a conclusion indicator; 6 of these (12% of the total) were particles like *hence*, and the other 3 (6% of the total) were verbs like *reveals that*. The total of passages with a premiss indicator and passages with a conclusion indicator is greater than 25 because one passage had two indicators, one of each type.

Table 4. Vertical complexity of a sample of 50 arguments and inferences			
Premiss supported	Implication drawn from conclusion	Both	Neither
2 (4%)	5 (10%)	0 (0%)	43 (86%)

Complexity of argument structure: Table 1 indicates the complexity of the structure of argument or inference to which the given passage belongs, on three dimensions: the number of stated premisses in the argument or inference under examination, the number of these premisses which are themselves supported by argument or inferred, whether there is a further inference from the conclusion of the argument or inference under examination. In the vast majority of cases (43, or 86%), the structure was extremely simple, with just a single inferential move. In only 2 cases (4%) was a premiss supported by argument, and in only 5 cases (10%) was a further conclusion drawn from the conclusion of the argument or inference under examination. In no case did both happen. As might be expected, the cases where a further conclusion was drawn came from proofs in mathematics (3, or 6%) and theoretical physics (1, or 2%); the remaining case, curiously enough, came from a piece of fiction, a short story. The vertical simplicity of the arguments and inferences in this sample may however be a function of the method of sampling, as I discuss below.

Table 5. Horizontal complexity of 50 arguments and inferences: number of premisses						
1	2	3	4	5	6	100
18 (36%)	11 (22%)	9 (18%)	5 (10%)	4 (8%)	2 (4%)	1 (2%)

Perhaps surprisingly, most of the arguments and inferences had more than one premiss. There were 18 one-premiss arguments (36%), 11 two-premiss arguments (22%), 9 three-premiss arguments (18%), 5 four-premiss arguments (10%), 4 five-premiss arguments (8%), 2 six-premiss arguments (12%), and one argument (2%) whose premisses consisted of 100 entries in a table. In general, arguments with 4 or more premisses involved rather loose reasoning from a mass of connected information about some particular situation–for example, the effects of various decisions by its founder on the early history of a private school in Canada, the effect of various changes in the supply of and demand for qualified young job applicants in Italy on youth unemployment, the effect of the introduction of various long-distance communication services on the frequency of social visiting by aboriginals in central Australia.

Argument schemes: An argument scheme (also known as an argumentation schema) is a pattern of reasoning which is valid (in any of the above-mentioned senses) under certain conditions, which can be articulated as answers to a set of critical questions about arguments of the scheme. Examples are modus ponens, inference to the best explanation, means-end practical reasoning, inference from a significant difference in outcome in a randomized trial to a causal relationship. I classified each passage as belonging to an argument scheme. The classification was difficult in many cases. I had no pre-conceived taxonomy of argument schemes, and am in fact skeptical of the view that all arguments and inferences can be classified in a pre-conceived taxonomy. I list below the argument schemes I found, in descending order of frequency, with the number of occurrences in parentheses after the name of schemes which occurred more than once:

- evaluation by criteria (11)
- means-end practical reasoning (5)
- existential generalization (4)
- interpretation of text(s) (4)
- direct conclusion from observations (3)
- mathematical deduction (3)
- cause-to-effect reasoning (2)
- generalization from instances (2)
- imputation of fallacy (2)
- qualitative description of non-statistical data (2)
- qualitative description of statistical data (2)
- argument by analogy
- application of definition to a case
- conceptual analysis
- classification by criteria
- inference to the best explanation
- inference to a possible explanation
- pros-and-cons decision-making
- pros-and-cons evaluation
- prediction from historical data
- repetition
- semantic entailment

(The total adds up to 51 rather than 50, because one argument was judged to exemplify two schemes at once.) Readers are invited to attempt their own classification, using a favoured taxonomy of argument schemes or simply their own intuitions.

One argument scheme which seemed to be used particularly poorly was the use of a quotation to support a claim about what its author believed. Of the four such arguments, classified under the label "interpretation of text(s)," three were invalid; the quoted statement simply did not support the claim attributed to its author on that basis. (The only valid example came from the use of a Biblical quotation in a 17th century sermon which led up to the witch hunts in Salem, Massachusetts.) If these few instances are at all representative of the use of quotations in scholarly writing to support claims about what people think, then there is room for improvement in educating university students to use this form of argument.

Discipline or genre: Most of the passages fell quite obviously into a definite discipline or literary genre. Most of the standard arts and science disciplines found in a contemporary university were represented; in alphabetical order, they included anthropology, art history, chemistry, economics, history (9), literary criticism, mathematics (2), philosophy (3), physics, political science (6), psychology, social work, sociology, statistics, and theology (2); of the standard arts and science disciplines, only biology, geography, linguistics, music history and religious studies are missing from this list. There were comparatively few passages from professional fields (one each from business, computer engineering, environmental engineering, museum curatorship, public administration, surgery and vocational guidance), perhaps because few publications in these fields are books. Genres represented included bibliographies, biographies (4), novels, short stories, sermons and travel memoirs.

Strength of inference link claimed: Only 9 of the passages (18%) incorporated some epistemic modal qualifier which indicated how strongly the author took the premisses to support the conclusion. Of these, 5 claimed a necessary link, typically with the qualifier *must*, 1 claimed probabilistic support, and 3 claimed what we might call "possibilistic" support with the verb *suggests that* or the qualifier *seems*. The claimed inferential strength bore little resemblance to the actual strength, as far as I could judge. Of the 5 passages whose authors claimed a necessary

link, 3 were invalid, 1 materially valid and 1 valid *ceteris paribus*. The passage claiming probabilistic support was valid *ceteris paribus*. Of the three passages claiming possibilistic support, 1 was formally valid, 1 was valid probabilistically, and 1 was invalid.

Linked and convergent structure with parallel premisses: Some contemporary textbooks in informal logic, and some theoretical publications (e.g. Freeman 1991, Snoeck Henkemans 1992, Vorobej 1994, Vorobej 1995a, Vorobej 1995b, Walton 1996, Snoeck Henkemans 2000), distinguish linked from convergent arguments. Both types have more than one premiss, each of which is offered in direct support of the conclusion. Intuitively, in linked arguments the premisses work jointly to support the conclusion, whereas in convergent arguments the premisses work independently. As Snoeck Henkemans (2000) shows in her recent review of this literature, there are various criteria for making the intuitive distinction precise. My own view is that the fundamental questions involved in applying any such distinction are evaluative, not analytical. That is, the fundamental questions concern the effect on the strength of the argument of finding no grounds to accept a given premiss, or of finding grounds to reject it. These questions underlie proposed tests for the type of support which involve eliminating a premiss from the argument or negating it. Further, as Vorobej points out in his (1995b), the answers to these questions can differ from one premiss to another, thus giving us "hybrid arguments" which are linked with respect to one premiss (its elimination or negation being fatal to the argument) but convergent with respect to another (its elimination or negation having no effect on the strength of the argument, or at worst weakening it). Further, there are in fact three possible answers to any such question, not two. Elimination (negation) of a premiss can have no effect on the strength of the argument, can weaken the argument but not make it invalid, or can invalidate it. (It can also strengthen it, in cases where the eliminated (negated) premiss is negatively relevant to the conclusion. Such cases are ignored when classifying multi-premissed arguments as linked or convergent.) These three alternatives correspond respectively to independent support by the rest of the premisses, cumulative support, and linked support. It is a moot point, incidentally, whether these distinctions make sense for an argument whose conclusion does not follow from the entire set of premisses offered in direct support of it.

Table 6. Support relations in 32 multi-premiss inferences

Effect on strength of support†	Of eliminating a premiss	Of negating a premiss
none	6 (18.8%)	3 (9.4%)
none, or weakened	3 (9.4%)	3 (9.4%)
none, or strengthened, or invalidated	-	1 (3.1%)
none or invalidated	3 (9.4%)	3 (9.4%)
weakened	9 (28.1%)	7 (21.8%)
weakened or invalidated	1 (3.1%)	2 (6.3%)
invalidated	10 (31.3%)	13 (40.6%)

†Disjunctive entries reflect different effects from eliminating (negating) different premisses.

Rather than attempting to classify multi-premissed arguments in the sample as linked or convergent, then, I asked two questions about them. What would be the effect on the strength of

the inference of eliminating a premiss? What would be the effect of negating a premiss? The answer to either question would be disjunctive if there would be one effect from eliminating (negating) one premiss and another from eliminating (negating) another. Using the abbreviations "0" for no effect on inference strength, "w" for weakening, "i" for invalidating and "s" for strengthening, and taking concatenation to indicate disjunction, I found the following incidence of effects of eliminating a premiss in the 32 multi-premissed arguments in the sample: 0 6 (18.8%), 0w 3 (9.4%), 0i 3 (9.4%), w 9 (28.1%), wi 1 (3.1%), i 10 (31.3%). The classification involved considerable judgement, especially about passages where I thought the inference was invalid even before eliminating a premiss; in these cases, I made a hypothetical judgement about what would be the effect on the strength of the inference if one supposed that the obvious covering generalization was acceptable. If one takes premiss elimination as the test of whether support by the premisses is independent, cumulative or linked, then the sample of 32 multi-premissed arguments included 6 (18.8%) with independent support, 9 (28.1%) with cumulative support, and 10 (31.3%) with linked support; the remaining 7 (21.9%) were some sort of hybrid.

Using the same abbreviations, I found the following incidence of effects of negating a premiss in the 32 multi-premissed arguments in the sample: 0 3 (9.4%), 0w 3 (9.4%), s0i 1 (3.1%), 0i 3 (9.4%), w 7 (21.9%), wi 2 (6.3%), i 13 (40.6%). As with the elimination test, I made a hypothetical judgement about the effect of negating a premiss in cases where I judged that the argument was invalid to begin with. If one takes premiss negation as the test of whether support by the premisses is independent, cumulative or linked, then the sample of 32 multi-premissed arguments included 3 (9.4%) with independent support, 7 (21.9%) with cumulative support, and 13 (40.6%) with linked support; the remaining 9 (28.1%) were some sort of hybrid. In most of the 32 passages, the effect of negating a premiss was the same as weakening it; where there was a difference, negating a premiss produced a stronger effect: invalidating instead of weakening or having no effect, weakening instead of having no effect, strengthening instead of having no effect. Considerable judgement was involved in determining the effect of eliminating or negating a premiss. Others may arrive at different judgements.

Argument or inference: Of the 50 passages, 45 (90%) were arguments and the remaining 5 (10%) were inferences. All the inferences were reported. That is, the authors of these five passages were not drawing an inference themselves, but were reporting an inference drawn by someone else. To put the point contrapositively, in every case where authors drew a conclusion in their own name, they did so in order to get their readers to draw this conclusion themselves.

Table 7. Direct vs. reported in a sample of 50 inferences and arguments

	Direct	Reported
Argument	44 (88%)	1 (2%)
Inference	0 (0%)	5 (10%)

Direct or reported: Of the 50 passages, 44 (88%) were direct discourse and 6 (12%) were reported discourse. As just stated, 5 of the 6 passages of reported discourse were inferences. The remaining passage reported an argument.

3. Discussion

How representative are these 50 passages? To what extent are we justified in generalizing the results obtained with this sample? What practical conclusions might we draw, with the help of such generalizations?

The question of representativeness can usefully be divided into two. To what extent are we justified in taking the sample of 50 arguments and inferences to be representative of the "universe" from which it was drawn, namely, the arguments and inferences in the English-language books listed in the catalogue of McMaster University library in the year 2000? To what extent are we justified in assuming that the distribution of characteristics of arguments and inferences in this universe is similar to its distribution in wider or different populations of arguments and inferences?

Representativeness: sample of universe: Two types of errors are possible in sampling from a universe, sampling errors and measurement errors. A sampling error would involve some bias in the way the sample was selected from the universe. For example, people who return an opinion survey mailed by their elected representative are a biased sample, because they have selected themselves into the sample; it is well known that people who agree with an elected official are much more likely to return a questionnaire to that official than are people who disagree. A measurement error would involve an incorrect observation of whether a given member of the sample had a given characteristic. For example, in an opinion survey, a person's response to a question can be influenced by the context in which the question is posed, the circumstances in which it is asked, defects of the question (e.g. vagueness, emotional slanting), or confusion or error about what the question means; such influences may lead the analyst of the response to attribute to the respondent an opinion different from the one actually held.

Possible sampling errors: The 50 starting-points from which we began to read, looking for an inference or argument, were randomly selected from the following block of texts in the McMaster library: the first 50 lines of each of the first 500 pages of each English-language book listed in the catalogue in the year 2000 (counting two lines as one if a page has two columns). Each line in this block of texts had an equal chance to every other line in the block of being in the sample of 50 starting-points. If a given percentage of lines in this block of text has a certain characteristic (say, beginning with a letter between a and m inclusive), we can calculate the 95% confidence interval, i.e. the interval around that percentage within which the percentage of starting lines in our sample will fall 19 times out of 20 such random selections. If the percentage is 50, for example, then the sample percentage will be between 36.14 and 63.86 19 times out of 20. Note that this is a rather large interval, which becomes only slightly smaller for other percentages; if the population percentage is 30 or 70, for example, then the sample percentage will be within 12.70 percentage points (in the first case between 17.30 and 42.70, in the second case between 57.30 and 82.70) of the population percentage 19 times out of 20. (The proportion of individuals with a certain characteristic in a sample of size N randomly chosen from a universe in which a proportion p_u of the individuals have the characteristic will be within $1.96[p_u(1 - p_u)/N]^{½}$ of p_u 19 times out of 20. See Blalock (1979, 214).) To conclude that the starting-points were representative of all lines in the corpus of English-language books in the McMaster library catalogue in a given respect, one has to assume that there is no difference in the given respect between these lines as a whole and the block of text from which the sample was actually selected. Equivalently, one assumes that there is no difference in the given respect between the block of text from which the sample was actually selected and the remaining texts which were excluded: pages after 500 in long books, lines after 50 in the first 500 pages of books

with a lot of lines on the page. In general, there seems no reason to suspect any systematic bias in excluding the last pages of long books or the last lines of long pages from possible starting-points.

A much more likely source of bias is the number of lines between a starting-point and the start of the first complete inference or argument. As we discovered, inferences and arguments are not uniformly distributed within books in McMaster's library. There are whole books which contain not a single argument or inference, e.g. census reports and bibliographies. There are others where one has to read for dozens of pages before coming across an inference or argument; in our sample, the list of Canadian railway statutes (Dorman 1938) and the checklist of Australian rare books published between 1788 and 1900 (in Wantrup 1987) are good examples. In other cases, e.g. mathematical proofs like those in Bolza (1931) and Butash & Davisson (1991), one can scarcely read for two lines before encountering a new argument. Let us use the term *argumentative (inferential) density* to refer to the frequency with which arguments (inferences) occur in a text. The huge differences in argumentative and inferential density which we encountered mean that our sample is not a random selection of arguments and inferences from the universe from which we randomly selected our starting-points. Consider two hypothetical arguments A and B in the universe. If we count back from the first line where argument A occurs, looking for the immediately preceding argument or inference, we discover that A's immediate predecessor starts 1,000 lines earlier in the universe. If we do the same for argument B, we discover that B's immediate predecessor starts two lines earlier in the universe. This means that argument A is 500 times more likely to be included in our sample of arguments and inferences than is argument B. The reader can check personally how unrepresentative our sample is in this respect by noting the disproportionately large number of passages taken from the very beginning of a book. Suppose that the average number of pages in a book in our universe is 250. Then we should expect roughly 2 passages to occur in the first 10 pages, since 10 pages is 4% of 250 and 2 is 4% of 50. In fact, we have 6 passages from the first 10 pages (passages 2, 7, 9, 11, 28 and 42), including three from the very first page (passages 9, 28 and 42). In each case, we selected the first argument or inference to occur in the book in question, and we started at the beginning of that book because there were no inferences or arguments between the starting-point selected in a previous book in the catalogue and the book where we found an argument or inference near the beginning; in some cases, we had to read through several short books in sequence before finding an argument or inference.

In order to apply confidence-interval estimates to distributions of a given variable in our sample of 50 arguments and inferences, one has to assume that there is no systematic relationship between that variable and the argumentative/inferential density of the text in which an argument or inference appears. Take for example the variable of validity status. One would have to assume that arguments and inferences in the universe whose immediate predecessor is 2 lines back have the same distribution of validity statuses (say, 3% formally valid, 6% semantically valid, etc.) as those with an immediate predecessor 3 lines back, 4 lines back, etc. For most of the variables discussed in the results section, this assumption is most probably false. Semantically valid arguments, for example, are much more likely to occur in mathematical proofs than in other types of arguments, but mathematical proofs occur in more argumentatively dense texts than most other arguments. Vertical complexity is by definition greater in argumentatively dense texts. Explicit verbal cues to the presence of argument or inference may be more likely in argumentatively dense texts, where vertical complexity imposes a requirement to clarify to the reader what assertion is being used to support what. The argument scheme of mathematical

deduction occurs more frequently in vertically complex arguments, which are necessarily found in argumentatively dense texts. Multi-premissed arguments of this scheme are more likely to be linked in structure (i.e. to be such that elimination or negation of a premiss invalidates the argument). Argumentatively dense texts are more likely to be direct than argumentatively sparse texts, and probably have a higher ratio of argument to inference than argumentatively sparse texts. Thus our sample of 50 arguments and inferences is likely to have under-represented the percentage of arguments and inferences in our universe which have the following characteristics:

- explicit verbal cues to the presence of argument or inference
- vertical complexity, with a premiss being supported by argument or a further conclusion drawn from the conclusion, or both
- an inference which is formally or semantically valid
- mathematical deduction
- linked structure
- an argument (as opposed to an inference)
- direct (as opposed to reported)

Without further investigation, one can only guess at how much the above-mentioned characteristics are under-represented in our sample.

The combination of a rather small sample, which would have a large margin of error if it had been randomly selected from our universe, and a biased selection process means that one can reasonably make at best extremely rough generalizations from the distribution of some variable in our sample to its distribution in our universe. One can have some confidence that characteristics which occur either almost always or almost never in our sample likewise occur either almost always or almost never, respectively, in the universe from which we selected our sample.

Measurement errors: We followed a careful procedure for identifying inferences and arguments, and for stating them in standard format using as far as possible the exact words of the original text. We had enough disagreements about identification, and even about standardization, that we recognize that someone replicating our task with the same selection of starting-points would come up with different passages (and even different standardizations) in some cases. It is doubtful that there would be many such cases, if the other person were well trained in argument identification and argument standardization and were following the definitions and criteria outlined above. I doubt too that any such differences would materially affect the distribution of characteristics of the passages in the sample.

The identification of characteristics of the passage (verbal cues if any to the presence of argument, validity status, number of premisses, etc.) must be regarded as preliminary. I did this myself, without checking by others. For some characteristics–such as verbal cues, number of premisses, field, or directness–inter-rater reliability is likely to be high. For others–such as validity status, type of argument or the effect of eliminating or negating a premiss–inter-rater reliability is likely to be low. For any of the variables, one needs to have the passages independently classified by well-trained raters in order to be sure of the classification.

Representativeness: Universe of broader or different populations: How similar are the arguments and inferences in the English-language books of McMaster University's library to the following:

- the arguments and inferences in the libraries of research universities?
- the arguments and inferences in the libraries of less research-intensive colleges and universities?
- the arguments and inferences which North American undergraduate students read as part of their undergraduate education?
- written arguments and inferences in English-language books generally?

- written arguments and inferences in English-language publications of all types?
- spoken arguments and inferences in English?
- arguments and inferences in languages other than English?

Without sampling like that reported in this paper, only educated guesses are possible in answer to these questions. McMaster University is a member of the Association of Research Libraries, to which at present more than 120 "libraries of North American research institutions" (www.arl.org/members.htm, 3 May 2001) belong. Its archives include a substantial collection of 18th century publications, as well as the papers and personal library of Bertrand Russell, along with ancillary material. It has a medical school and faculties of business and engineering, but no law school or education faculty. It is a Canadian university. These facts mean that its collection is reasonably substantial, but more sparse in education and law than collections of research institutions with a faculty of education or a law school. Its collection undoubtedly includes proportionately more material about Canada and its provinces than those of American research universities; it probably includes more material about Commonwealth countries, who are the subject of a considerable number of the passages in our sample. As a result, the collection probably includes proportionately less material about the United States and its 50 states than those of American research universities; even on the most generous construal, only 7 of our 50 passages deal specifically and exclusively with the United States. It is hard to tell whether these differences in subject-matter of the McMaster collection would be reflected in characteristics of its arguments and inferences.

Looking over our sample of 50 inferences and arguments, the reader can quickly recognize that few of them would be the normal reading fare of a typical college undergraduate in North America. It is one thing for a book to be in a university library. It is another thing for an undergraduate actually to read it. The books undergraduates read are likely to be easier to read, less technical and more general in subject-matter than many of the passages in our sample. The characteristics of the arguments and inferences in the reading of college undergraduates in North America are likely to differ correspondingly from the characteristics of those in the libraries of the colleges or universities they are attending. It seems safe to conclude that there would be proportionately fewer mathematical deductions than appeared in our sample, and proportionately fewer bibliographies, census reports or publications more than 50 years old. If it were possible, one could get a sample more representative of the reading matter of college undergraduates by selecting from a universe of borrowings from a college or university library, since there is a close (not exact) correspondence between materials borrowed and materials read. If one wants to use the results of this sort of sampling to provide some sort of guidance as to what to focus on in an undergraduate course in critical thinking, a sample of student borrowings would be more useful than the sample we selected.

I have no reason to conclude that arguments and inferences in English-language scholarly and scientific periodicals differ in the distribution of their characteristics from those in English-language scholarly and scientific books. Since the authors are the same and the intended audience is generally the same, there is unlikely to be any systematic difference. We excluded periodicals from our universe because a periodical has far more pages than a book; to get a random sampling of starting-points from a universe which included periodicals as well as books, we would have had to count each volume of each periodical as a separate book, which was impractical.

Similarly, I have no reason for thinking that scholarly and scientific publications in languages other than English differ in their argumentative characteristics from those published in English. On the other hand, I have no strong evidence that they are the same.

Written material which is not scholarly or scientific (e.g. news reports, feature articles, opinion columns, editorials and letters to the editor in newspapers and magazines) are in general less technical and less formal than scholarly and scientific writing. One would expect fewer mathematical deductions in such material, fewer semantically valid arguments, and more need to supplement stated premisses from the context in order to apply the test for validity described at the beginning of this paper.

Arguments and inferences which are not written but are spoken, signed or thought are obviously more likely to need supplementation from the context of their utterance in order to be understood, let alone to be evaluated.

4. Conclusion

Despite all the difficulties discussed above, there is I think some value in this first attempt to use random sampling methods to see what human arguments and inferences are actually like. In particular, I have shown that a clearly stated test for the validity of an inference can be applied successfully to a sample of arguments selected by random methods and coming from a wide variety of fields. In all but one of the 50 inferences to which I applied the test, there was the required overlap of subject-matter between premiss(es) and conclusion, and a sensible judgement could be reached as to whether the conclusion followed, according to this general test, and if so whether definitely, probabilistically, *ceteris paribus* or merely as a possibility. (The one exception is in my opinion a rather odd argument.) It also should be clear from this sample that formally valid arguments and inferences are rather rare, even in the scholarly and scientific contexts where we might expect to find them more often. Thus, if formal logic applies to real-life arguments and inferences, it must in most cases do so indirectly rather than directly. It should also be clear from this sample that the evaluation of inferences in real-life reasoning requires substantive knowledge in many cases, often specific to the field to which the reasoning belongs.

References

Blalock, Hubert M. 1979. *Social Statistics*, revised second edition. New York: McGraw-Hill.

Freeman, James B. (1991). Dialectics and the Macrostructure of Arguments: A Theory of Argument Structure. Englewood Cliffs, NJ: Prentice-Hall.

Hitchcock, David (1985). “Enthymematic arguments”, *Informal Logic* vii, 83-97.

Hitchcock, David (1992). “Reasoning by analogy: a general theory”, in Stephen P. Norris (ed.), *The Generalizability of Critical Thinking: Multiple Perspectives on an Educational Ideal* (New York: Teachers College Press), 109-124.

Hitchcock, David (1994). “Validity in conductive arguments”, in Ralph H. Johnson and J. Anthony Blair (eds.), *New Essays in Informal Logic* (Windsor, ON: Informal Logic), 58-66.

Hitchcock, David (1998).“Does the traditional treatment of enthymemes rest on a mistake?” *Argumentation* 12, 15-37.

Snoek Henkemans, A. Francisca (1992). Analyzing Complex Argumentation. The Reconstruction of Multiple and Coordinatively Compound Argumentation in a Critical Discussion. Amsterdam: Sicsat.

Snoek Henkemans, A. Francisca (2000). “State-of-the-art: The structure of argumentation,” *Argumentation* 14, 447-473.

Vorobej, Mark (1994). "The TRUE test of linkage," *Informal Logic* 16, 147-157.
Vorobej, Mark (1995a). "Linked arguments and the validity requirement," *Argumentation* 9, 291-304.
Vorobej, Mark (1994). "Hybrid arguments," *Informal Logic* 17, 289-296.
Walton, Douglas (1996). *Argument Structure: A Pragmatic Theory*. Toronto: University of Toronto Press.