OSSA Conference Archive   OSSA 2

May 15th, 9:00 AM - May 17th, 5:00 PM

Claim Strength and Burden of Proof

Jeremy Bailenson
Northwestern University

Lance J. Rips
Northwestern University

BURDEN OF PROOF IN CONVERSATIONAL ARGUMENTS[1]

Jeremy N. Bailenson
Lance J. Rips
Department of Psychology
Northwestern University

Abstract:

In this paper, we report results from experiments in which people read conversational arguments and then judge (a) the convincingness of each claim and (b) the individual speakers' burden of proof. The results showed an "anti-primacy" effect: People judge the speaker who makes the first claim as having greater burden of proof. This effect persists even when each speaker's claims are rated equally convincing. We also find that people rate claims less convincing when they appear in the first part of an argument than when they appear in isolation.

***

Disputes are a common occurrence in everyday discourse and typically involve two or more individuals attempting to convince each other or a third party of a particular claim. This paper will primarily be concerned with two-sided arguments, as in the following dialogue from a case on Internet censorship (ALA v. Pataki):

(1)a. MR. HERSHLER: Plaintiffs have not sued the district attorneys in this action. They have only sued Pataki—who is not going to be prosecuting anybody—and the Attorney General's office.
b. THE COURT: That is not the issue right here.
c. MR. HERSHLER: I respectfully disagree, your Honor.
d. THE COURT: Have you finished this line of questioning with this witness?
e. MR. HERSHLER: I am going to be very brief, your Honor, I can assure you.
f. THE COURT: I know that. The question is: Are we going to spend much time on what this witness understands the Act to mean?
g. MR. HERSHLER: Well, his understanding of the Act will be a current running through his testimony.
h. THE COURT: Move on, please.

When people in an argument such as (1) make claims, they usually try to defend those claims with accepted evidence. We can view this evidence as an attempt to reduce an initial burden of proof that their opening positions set. If one speaker presents an initial claim that is far more convincing than the other speaker's, then that individual has little to do to prove his or her position. The burden would lie on the other participant, who would then be forced to overcome the disparity between the quality of the two positions. In this sense, the burden of proof shifts from one speaker to the other as the argument progresses, depending on certain rules.

In the United States's judicial system, judges use formalized rules of burden of proof to decide whether information is admissible as evidence, and jurors use them in order to assess the quality of the evidence (Wigmore, 1937). The precepts dictating burden of proof, however, are not as well defined in disputes that occur outside a trial setting. In this paper we report results from an experiment which looks at burden of proof

and claim strength in informal arguments. Subjects determined burden of proof in two types of arguments: *two-sided arguments* that occur in dialogues and *one-sided arguments* where a single speaker makes claims in support of a position. In addition, subjects rated the strength of the same claims when they appeared in the two-sided arguments, in the one-sided arguments, and in *isolation* from the argument context. Before describing the experiment, however, we review some previous research on what makes claims and arguments convincing.

Local versus Global Strength

During an argumentative dialogue, speakers can choose among a number of acceptable conversational moves (Orsolini, 1993). Each move can produce a unique result, changing the direction of the conversation. We can think of participants and judges "keeping score" in an argument (Lewis, 1979), with utterances earning points depending on a variety of factors—the history of the conversation, the presuppositions of the arguers, and the degree of cooperation they establish.

Arguers use cue words and phrases, such as *in the first place* and *so*, to mold expectations of forthcoming claims and to highlight the inference relations between claims within the argument structure: "The discourse system analyzes such support moves by identifying some principle of support, and setting up mappings between this principle, the stated fact of support, and the claim being supported" (Reichman-Adar, 1984, p. 177). Thus, in discourse-based arguments, participants and judges extract information, not only from the meanings of individual claims, but also from the role of those claims in the overall dialogue.

This dialogue context has a direct impact on how people perceive burden of proof. A speaker's burden clearly depends on the strength of his or her individual claims, and previous researchers have defined claim strength in many different ways ranging from number of words in the claim (Gulley & Berlo, 1956) to how "bold" it is (Baron, 1991). However, claim strength also depends on the relation of the claim to *other* statements. In this paper we would like to draw the distinction between *local* and *global* strength. Local strength concerns the inherent characteristics of individual claims that make them convincing. The local strength of a claim is the same regardless of its context. Global strength is a measure of how well each claim fits into the overall context of the argument: how relevant it is to the initial context, how responsive it is to previous claims, and how well it supports the speaker's position as a whole. Intuitively, burden of proof should depend on both these measures of strength.

Order of Claims

Although people who judge disputes clearly pay attention to the local strength of claims (e.g., Bailenson & Rips, 1996), research shows that they also attend to structural factors. Presenting two groups of identical claims in different orders can produce separate patterns of judgment (McGuire, 1957). For example, some studies find a recency effect, where the information from the end of an argument is remembered better and is more persuasive than information from the beginning. Recency effects appear in situations where subjects read two contradictory messages (Crano, 1977), in mock trials where subjects judge guilt or innocence (Furnham, 1986), and in instances where subjects evaluate traits of other people (Luchins & Luchins, 1984).

Researchers also find ample instances of primacy effects in judgment. In these experiments, people retain more information from the beginning of an argument, and this information then biases how later claims are processed.

Pennington and Hastie (1986) showed this phenomenon in mock jurors: According to their theory, the first bit of information the jurors received initiated a mental story about the supposed crime, thus framing how they interpreted later claims. Primacy effects also occur in judgments of how well other people perform a task (Benassi, 1982).

Evidence of primacy and recency usually comes from studies that interchange the positions of claims or evidence in unorganized serial lists. Global strength may also vary, however, as a result of conversational structure. Evidence for this hypothesis comes from a study by Bailenson and Rips (1996) in which subjects read naturalistic arguments between two speakers. The results showed an *anti-primacy* effect, in which the speaker who initiates the debate incurs more burden of proof. This bias against the first speaker seems to be due to the structured dialogue, since previous studies have not shown a disadvantage in persuasiveness for items at the top of a serial list.

This effect may depend on differences in the roles of the two speakers in the argument. The first speaker or *initiator* stages the argument, suggesting an underlying structure or script molding the remainder of the argument (Scholtens, 1991). The second speaker or *recipient* does not have as much control in molding the hierarchy of the conversation, as his or her job is directed towards following the text and understanding: collecting information, signalling irrelevant ideas, pointing out obscurities and falsities, and giving general reactions. Although both speakers can employ strategies throughout the argument to change its structure (Walton & Krabbe, 1996; Resnick et al., 1993), the initial division of roles may help explain the heavier burden of proof for the first speaker.

Anti-Primacy and Convincingness in Arguments

We attempted to find out whether the anti-primacy effect would occur in arguments in which the claims of each speaker were equally convincing (had the same local strength). To accomplish this, we manipulated which side of an argument was taken by the first or second speaker. For example, in version A of an argument the first speaker argues in favor of abortion while the second speaker argues against it. In version B the first speaker argues against abortion, and the second speaker argues in favor of it. However, the speakers use exactly the some claims, pro and con, in each version. In this set up, it is impossible to argue that anti-primacy is due to the local strength of the claims, since each version contains the same claims. We predicted that anti-primacy would persist even though the claims are identical, because the first speaker's role as initiator sets the context for the argument.

This experiment also allows us to test a second hypothesis. People's impression of the (global) strength of an individual claim should depend on its position within an argument. Claims that appear early in an argument are subject to critical scrutiny as the dialogue continues. Hence, early claims, like first speakers, may be at a disadvantage within a debate. To find out if this is so, we compared subjects' ratings of a claim's strength in the context of an argument with ratings of the strength of the same claim in isolation. Thus, the experiment aims to detect effects of argument structure both on burden of proof and on the perceived strength of individual claims.

We first describe how we designed the arguments we gave to subjects; then we outline our procedure for presenting the arguments to subjects; and finally, we examine the data for relations among argument structure, burden of proof, and global claim strength.

The Arguments. We constructed booklets containing nine different arguments with six statements apiece. Each argument appeared on a separate page. The content of the arguments varied, but they typically involved entertainment, politics, or college life, as example (2) illustrates:

(2)a. Ted: Abortions should be legal.
b. Ron: Abortions should not be legal.
c. Ted: The right to have an abortion for a woman is fundamental.
d. Ron: The rights of the unborn child are more important than the short term convenience of the mother.
e. Ted: The freedom to make decisions about your own body is one of the most important rights a person can have.
f. Ron: When an abortion is performed, the infant dies—murder is not a right.

Each argument appeared in two versions, with the pro side beginning in one version and the con side in the other. We produced the second version of each argument by reversing the positions of claims a and b (*initial claims*), claims c and d (*middle claims*), and claims e and f (*final claims*). For the sample argument (2) above, we changed the orders of the claims to produce the second version in (3):

(3)a. Ron: Abortions should not be legal.
b. Ted: Abortions should be legal.
c. Ron: The rights of the unborn child are more important than the short term convenience of the mother.
d. Ted: The right to have an abortion for a woman is fundamental.
e. Ron: When an abortion is performed, the infant dies—murder is not a right.
f. Ted: The freedom to make decisions about your own body is one of the most important rights a person can have.

We created the arguments around the nine topics and then changed the order of the claims as we have just illustrated to determine which side began the argument.

We were also interested in comparing burden of proof and claim strength in non-conversational, one-sided arguments consisting of claims made by one speaker supporting a position. We therefore created another set of arguments by splitting the nine two-sided arguments into 18 one-sided ones. For example, two-sided argument (3) was separated into one-sided arguments (4) and (5):

(4)a. Ron: Abortions should not be legal.
b. Ron: The rights of the unborn child are more important than the short term convenience of the mother.
c. Ron: When an abortion is performed, the infant dies—murder is not a right.

(5)a. Ted: Abortions should be legal.
b. Ted: The right to have an abortion for a woman is fundamental.
c. Ted: The freedom to make decisions about your own body is one of the most important rights a person can have.

The Experiment. There were three groups of subjects in this experiment: One group read two-sided arguments (e.g., argument (2) or (3)), a second group read one-sided arguments (e.g., (4) or (5)), and a third group read randomly ordered claims from the same arguments. Subjects in the one-sided and two-sided conditions received two booklets, each booklet containing all the argument topics. In one booklet, they rated the strength of each claim in the argument, and in the other they assessed burden of proof. Subjects in the isolated-claims condition received just one booklet containing the claims from all nine topics, shuffled together so that the argument structure would not be evident. These subjects rated the strength of each claim.

In the two-sided condition, subjects chose the speaker with more burden of proof by answering the question, "Which of the two speakers has got more to do in order to prove [he or she is] correct?" They indicated their answer by circling the name of the speaker and then rating their confidence in the answer on a 7-point scale. Subjects assessed the strength of the individual claims in one of two ways. Half the subjects in this condition rated strength by answering the question, "How convincing is the claim, in and of itself?". The remaining subjects rated strength by answering the question, "How well does the claim fit into the argument as a whole?". The strength question appeared underneath every claim in the argument, and subjects responded by circling a number on a 7-point scale. Each argument appeared on a separate page of the booklets, and an individual subject rated only one version of each of the nine arguments (e.g., argument (2) or (3), but not both). The order of the two booklets was balanced, so that half the subjects judged burden of proof first and half rated strength first.

Subjects in the one-sided condition followed much the same procedure as those in the two-sided condition. However, these subjects read and rated only one side of each of the nine arguments (e.g., argument (4) or (5), both not both). In this condition, subjects judged burden of proof by rating, "How much work does Ron [Ted] have to do in order to prove he is correct?", and they rated claim strength in the same way as the subjects in the two-sided condition.

In the isolated-claims condition, subjects rated the convincingness of each claim. We presented the claims to them in one of four randomized orders, with the argument topics dispersed in the list. For example, a claim about abortion could appear next to a claim about baseball. Subjects were told to make each rating independent of the others.

There were 23 subjects in the isolated condition, and 48 subjects each in the one-sided and two-sided conditions.

Claim Strength and Burden of Proof

In examining the results of the experiment, we look first at subjects' judgments of burden of proof and then at their ratings of claim strength.

Anti-primacy effects. Across conditions, subjects chose the first speaker as having more burden of proof in 58% of the arguments. This difference was reliably different from chance according to standard statistical tests[2], consistent with our earlier findings. Anti-primacy persisted even though there was no reliable difference between the strength of claims offered by the first speaker (mean strength rating = 3.96) and the claims offered by the second speaker (3.80) when these claims appeared in isolation. So anti-primacy is not due solely to the strength of specific claims, but depends on the roles of the speakers in the argument.

Claim strength. By having different groups of subjects rate the same claims when they occurred in isolation, in one-sided arguments, and in two-sided arguments, we can examine the way the argument structure affects subjects' perception of the claims' strength. To aid this comparison, we present in Figure 1 the mean strength ratings for the claims in these contexts. Although separate subjects rated the convincingness and the relevance of the claims, there was actually no difference between these types of ratings (mean convincingness = 3.98, mean relevance = 3.95). So, in Figure 1 and in the remainder of the paper, we report as strength ratings the averages of these two measures.
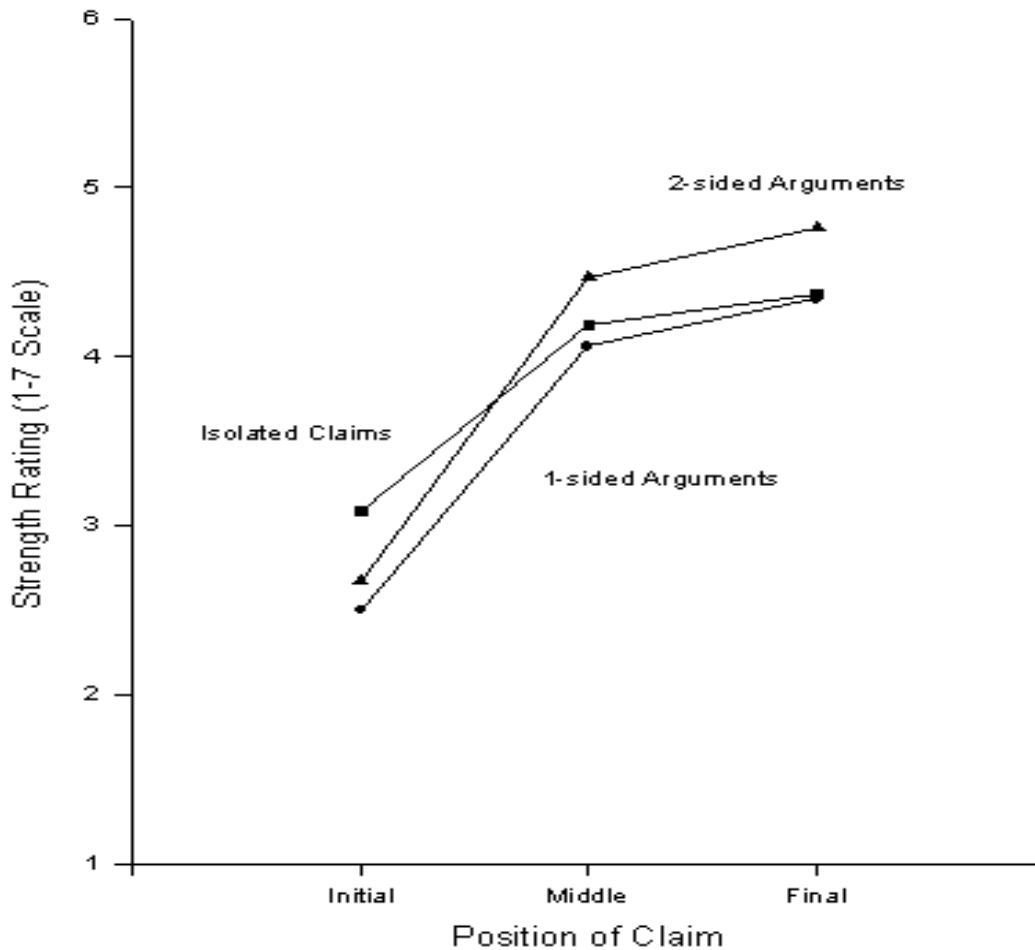

Figure 1 shows an overall increase in strength from initial claims to middle claims to final claims. This increase appears even when the claims appear in isolation and is therefore partly due to differences among the claims themselves—to variations in their local strength. However, the figure also shows that argument structure modulates this increase. Initial claims received reliably higher ratings in isolation than in the arguments (the mean strength is 3.09 for claims in isolation versus 2.50 in one-sided and 2.67 in two-sided argument).[3] The initial claims are the points at issue in the rest of the argument, and the subsequent discussion apparently raises questions for the subjects about their convincingness. It is noteworthy that this decrease occurs even when the initial claim appears in one-sided arguments, since in these arguments the other claims provide support for it. Claims in the middle position and claims in the final position were rated as slightly stronger in the two-sided arguments than in isolation or in one-sided arguments, but this difference was not statistically significant.

Correlation between strength ratings and burden of proof. Although the overall strength of the claims is the same for the pro and con sides of the arguments, strength of claims does vary from one subject to the next. By comparing the subjects' strength ratings for individual claims with their burden-of-proof choices, we can gauge how well the strength of initial, middle, and final claims predicts burden of proof. For the two-sided arguments, we first computed the difference between the strength of the initial claim on the pro side and the strength of the initial claim on the con side (e.g., strength of claim (2a) — strength of (2b)), and we did the same for the middle and the final claims. We then separately correlated these three differences with the subject's choice of which speaker (pro or con) had greater burden of proof.[4] For one-sided arguments, we simply correlated a subject's strength rating for initial, middle, and final claims with his or her rating of how much burden the corresponding speaker had.

These correlations between strength and burden of proof should be negative, since the greater the strength of the pro side claims, the smaller the pro side's burden. In fact, all correlations were in the correct direction, ranging from -.20 to -.45. In displaying these correlations in Figure 2, however, we have dropped the minus signs, so that the stronger the relation between strength and burden, the larger the value on the y-axis.
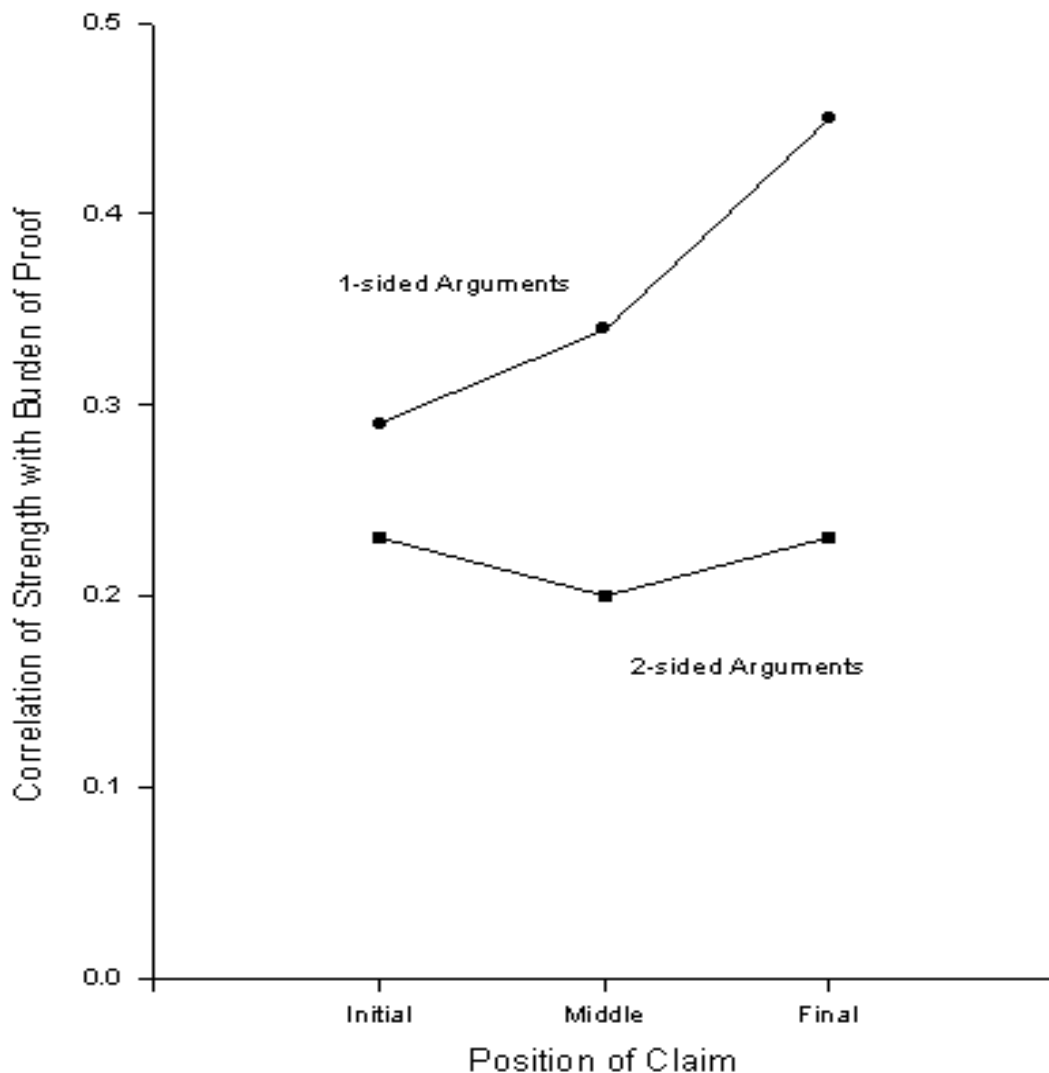
The correlations between strength and burden of proof in the two-sided arguments were not statistically significant, and they varied little across the three argument positions. For the one-sided arguments, however, claim strength correlates reliably with burden of proof in middle and final positions. One-sided arguments contain fewer structural characteristics than two-sided arguments, and as a result, subjects may have relied more on the strength of the individual claims to evaluate the speaker's burden.

Conclusions:

The results of the experiment show that the first speaker is typically assigned more burden of proof than the second speaker, even when the claims of the two speakers are equally convincing. The subjects were not sensitive to the difference between how convincing a claim is and how relevant it is to the argument as a whole. Nevertheless, placing a particular claim in the context of an argument affects subjects' perception of its strength. Claims made early in an argument received lower strength ratings than when they appeared in isolation. In addition, strength of claims located later in one-sided arguments tend to correlate more with burden of proof than do the early ones.

When evaluating arguments in discourse, people cannot entirely dissociate the claims from the overall configuration of the dialogue. Even if the initiators of arguments ensure that their claims are at least as persuasive as those of their opponents, they still may be punished with burden of proof. Likewise, claims that would normally persuade people when removed from the context of a conversational argument may not be as convincing when they appear in the initial position of a dispute. These findings call into question studies on persuasiveness that ignore the argument structure, and they indicate that argumentative conversations may have unique properties.

Notes

1. This research was supported by NSF Grant Number SBR-9514491. Correspondence about this paper should be addressed to Jeremy Bailenson or Lance Rips, Psychology Department, Northwestern University, 2029 Sheridan Road, Evanston, IL 60208.

2. $t(47)=3.28, p<.01$

3. $t(8)=3.25, p<.01$

4. For these purposes, we scored the choice +1 if the speaker chose the pro side as having more burden of proof and -1 if the speaker chose the con side. We then multiplied these scores by the subjects' confidence rating to get a measure of the degree to which the pro or con sided had the burden.

References

Bailenson, J. & Rips, L. J. (1996). Informal reasoning and burden of proof. *Applied Cognitive Psychology*, 10, 3-16.

Baron, J. (1991). Beliefs about thinking. In J. F. Voss, D. N. Perkins, & J. W. Segal (Eds.), *Informal reasoning and education* (pp. 169-186). Hillsdale, NJ: Erlbaum.

Benassi, M. (1982). Effects of order presentation, primacy, and physical attractiveness on attributions of ability. *Journal of Personality and Social Psychology*, 46, 1230-1240.

Crano, W. D. (1977). Primacy versus recency in retention of information and opinion change. *Journal of Social Psychology*, 101, 87-96.

Furnham, A. (1986). The robustness of the recency effect: Studies using legal evidence. *Journal of General Psychology*, 113, 351-357.

Gulley, H. E. & Berlo, D. K. (1956). Effect of intercellular and intracellular speech structure on attitude change and learning. *Speech Monographs*, 29, 288-297.

Lewis, D. (1979). Score keeping in a language game. *Journal of Philosophical Logic*, 8, 339-359.

Luchins, A. S., & Luchins, E. H. (1984). Conceptions of personality and order effects in forming impressions. *Journal of General Psychology*, 110, (165-196).

McGuire, C. I. (1957). Order of presentation as a factor in "conditioning" persuasiveness. In C. I. Hovland (Ed.), *The order of presentation in persuasion* (pp. 158-169). New Haven, CT: Yale Univ. Press.

Orsolini, M. (1993). "Dwarfs do not shoot": An analysis of children's justifications. *Cognition and Instruction*, 11, 281-297.

Pennington, N., & Hastie, R. (1986). Evidence evaluation in complex decision making. *Journal of Personality and Social Psychology*, 51, 242-258.

Reichman-Adar, R. (1984). Extended person-machine interface. *Artificial Intelligence*, 22, 157- 218.

Resnick, L. B., Salmon, M., Zeitz, C. M., Haley Wathen, S., Holowchak, M. (1993). Reasoning in conversation. *Cognition and Instruction*, 11, 347-364.

Scholtens, A. (1991). Planning in ordinary conversation. *Journal of Pragmatics*, 16, 31-58.

Walton, D. N. & Krabbe, E. C. (1995). *Commitment in dialogue*. Albany, NY: SUNY Press.

Wigmore, J. H. (1935). *A students' textbook of the law of evidence*. Brooklyn: The Foundation Press.