APPLIED COGNITIVE PSYCHOLOGY, VOL. 10, S3–S16 (1996)

Informal Reasoning and Burden of Proof

JEREMY N. BAILENSON and LANCE J. RIPS
Northwestern University, USA

SUMMARY

Informal arguments occur in casual discourse and typically feature participants trying to convince each other (or a third party) of their positions. As an argument progresses, one participant can accrue more burden of proof—will have to do more to prove he or she is correct. Some past studies have shown that evidence presented first is more effective than later evidence (primacy), while others have found that evidence presented last is more effective (recency). These studies largely focused on lists of evidence, however, and not on structured, persuasive dialogues. In the current experiments, subjects read dialogues between two speakers having a dispute. They then chose the participant with the greater burden of proof. The results demonstrated that burden of proof increases for the participant who offers the first claim in an argument (*anti-primacy*) and sometimes decreases for the participant who offers a final challenge (e.g., *What's your evidence?*). These results suggest strategies that can help participants gain advantages in disputes and that may assist judges in avoiding bias while evaluating arguments.

INTRODUCTION

Disputes typically involve two or more individuals attempting to convince each other or a third party of a particular position or claim. The disputes can occur in a formal arena, as in a planned debate, or in an informal context, as in discussions among family members and friends. In the present paper, we focus on disputes in which the participants appeal to reasons to support their positions. We ignore verbal conflicts that are primarily emotive.

Observers of the dispute often make judgments as to which of the opposing sides is doing a better job arguing his or her position. The outcomes of these decisions can have practical consequences: The audience's evaluation of Presidential debates, the judge's or jury's evaluation of arguments in criminal or civil cases, the scientific establishment's evaluation of competing theories, the negotiator's evaluation of rival arguments in a labour dispute, and the CEO's evaluation of competing strategies proposed by subordinates can all imply substantial gains or losses for participants and for society. The potential benefits and drawbacks from these decisions make it important to examine whether there are general factors that influence observers' judgments about which side made the more convincing case. Such factors span different dimensions of the argument, such as characteristics of the disputants (e.g.,

Address correspondence to: Lance Rips, Psychology Department, Northwestern University, 2029 Sheridan Road, Evanston IL 60208, USA.

We thank Sarah Brem, Norman Eliaser, Art Graesser, and Mike Shum for their helpful comments on an earlier draft of this article.

CCC 0888–4080/96/SI0003–14

Accepted 2 July 1996

Chaiken, 1987; Hovland and Weiss, 1951; Kelman and Hovland, 1953), and characteristics of the dispute (see McGuire, 1985, for a review). The present research addresses the discourse structure of a dispute and the ways this structure determines which participant has the burden of proof.

Burden of proof

When people who are arguing make claims, they often try to defend those claims with accepted evidence. In the United States judicial system, rules governing burden of proof determine what outcomes are possible on the basis of the evidence. Legal rules dictate, for example, that an individual who has been absent for seven years and who has not been heard from during that time is presumed dead. The burden of proof therefore rests on the party trying to prove the person is alive. Although there are several meanings of *burden of proof* in law, the associated rules generally determine a default decision when no further evidence on an issue is forthcoming (Wigmore, 1937). A related sense of burden of proof is common in academic, political and other controversies (Gaskins, 1992): The side of an issue that has the burden of proof is the side that will lose by default unless further evidence appears supporting that side. In this sense (and in some of its legal senses), burden of proof can shift during the course of a dispute. Once one side has fulfilled its burden by presenting convincing evidence, the burden will transfer to the other side of the debate, forcing that side to produce its own evidence or face the prospect of losing the controversy. It is this informal meaning of burden of proof that will concern us in this paper.

Argument structure and burden of proof

A typical dispute consists of an exchange of statements by two or more participants. For the purposes of this paper, we will call this group of statements an *argument*. Within such an argument, some of the participants' statements are *claims*—assertions they seek to establish (Toulmin, 1958). The claims made by a participant partially constitute his or her position in the argument. Arguments, however, are not simply lists of claims. Participants often try to provide a reason for a claim by introducing another that lends the first support. They also challenge the support for others' claims by asking questions like *What makes you say that?* or *How do you know?*, and they attack others' claims by offering direct refutations. In this way, arguments can become complex, structured discourse entities, consisting of many interrelated conversational moves: assertions, concessions, challenges, justifications, attacks and counterattacks. The argument often contains discourse markers (e.g., *so, because, thus*) indicating the relations among the claims (e.g., Reichman-Adar, 1984; Schiffrin, 1987).

The conversational structure of an argument also helps determine which of its participants has the burden of proof (Rips, in preparation). For example, if a participant concedes a claim to his or her opponent, the burden on the opponent normally becomes lighter. Conceded claims are mutually accepted and become part of common ground (Clark and Schaefer, 1989). This implies that the participant who originally made the claim no longer has to defend it and can proceed to use it to support further assertions. By contrast, the more claims a participant makes that are still under contention, the heavier the participant's burden. Evidence for this relation

between concessions and burden of proof comes from an experiment in which subjects read argumentative dialogues and chose which of the dialogue's participants had the burden of proof at the end of the passage (Rips, in preparation). The results showed that, as expected, the greater the number of a participant's claims that are mutually accepted, the lighter that participant's perceived burden.[1]

In addition to the effect of concessions, however, our earlier study unexpectedly found that the burden of proof was greater for the first speaker in the dialogue than for the second speaker over all argument types. We refer to this effect as *anti-primacy*, because it places the first speaker at a potential disadvantage during the course of the argument. It is the nature of the anti-primacy effect that we explore in this paper. Our interest in anti-primacy stems from the fact that it appears to contrast with a *primacy* effect sometimes reported in the persuasion literature. The primacy effect is the tendency of initially presented evidence to be more persuasive than later evidence (see Hogarth and Einhorn, 1992, for a synthesis of earlier research). Since more persuasive evidence should in turn reduce the burden of proof for the person who provides it, the primacy effect would seem to predict a lighter burden on the first speaker, contrary to our findings.

Primacy versus anti-primacy

There are a number of factors that could account for this seeming conflict between primacy in earlier research and our own anti-primacy effect. One such factor is that the arguments in our earlier experiment always began with a claim by the first speaker (e.g., *The El is the best form of transportation in Chicago*) followed by a question or challenge from the second speaker (*Why do you say that?*). It seems possible that the challenge focused the burden of proof on the first speaker to justify the expressed position. The challenge operates at a local level by calling attention to (or placing in doubt) the first speaker's claim.

It is possible, however, that anti-primacy has a more global cause that lies in the function that the first claim performs. The materials that investigators used in the studies that obtained primacy effects were typically lists of descriptions or traits. For example, the experimenter may have asked subjects to evaluate the overall likableness of a person on the basis of a series of trait adjectives, each of which may be positive or negative (e.g., Anderson, 1981). The weight subjects place on the first adjective then determines the presence or absence of primacy. In our study, however, the characters in the argument, rather than the experimenter, raised the issue under debate and took a stand on the issue. The first speaker, in particular, set the agenda for the argument by stating his or her opinion on a topic. By taking the issue-setting stand, the first speaker may have assumed the burden of proof and, in this way, become vulnerable to anti-primacy. This possibility is related to earlier findings on text comprehension that demonstrate the importance of an initial topic sentence (e.g., Kieras, 1978).

The factors we have just discussed certainly do not exhaust the possible reasons why anti-primacy occurred in our subjects' judgments, but they give us a place to

[1]For clarity in this paper, we refer to the characters in an argumentative dialogue as the *participants* or *speakers*, and we refer to people who took part in the experiments as *subjects*.

start. Both suggest that the conversational structure of disputes plays a crucial role in determining burden of proof and that conversational structure is not reducible to the simple linear order in which the evidence appears. The first experiment in the present paper varies whether or not a challenge occurs in the argument in order to see whether this changes the perceived burden of proof. We also vary whether the first speaker makes the first substantive claim or makes a neutral opening comment. If the perceived burden depends on an agenda-setting stand, then the first speaker should have greater burden only when he or she begins with a claim. In Experiment 2 we use a forced-choice paradigm to evaluate residual effects of challenges.

EXPERIMENT 1: BURDEN OF PROOF AS A FUNCTION OF ARGUMENT STRUCTURE

On each trial in this experiment, subjects read an argument between two speakers and then choose the speaker with the greater burden of proof. We compare three types of arguments to find out whether challenges produce anti-primacy (greater burden of proof on the first speaker). In two of the argument types, one speaker challenges the second by asking *What is your evidence for that statement?* This challenge occurs near the beginning in one of the arguments and at the end in the other. The third, control argument omits the challenge entirely, so that the dialogue consists only of assertions. If challenges increase the perceived burden, subjects should identify the challenged speaker as having a greater burden in the first two arguments relative to the control. The arguments also differed according to whether the first speaker begins with a substantive claim (e.g., *Abortion should not be legal*) or with a neutral greeting. If burden of proof depends on which speaker takes the issue-setting stand, we would expect the first speaker to have the greater burden when he or she starts with a claim. Finally, we varied which speaker had stronger claims, since claim strength could also determine who has the burden of proof.

Method

We constructed booklets of 12 different arguments, one argument per page. The content of the arguments varied, but typically involved political issues or college life. Subjects read the arguments one at a time, picked the speaker with greater burden of proof, and rated their confidence in this decision.

Argument types and challenges

The arguments appeared in three structural formats that we generated on the basis of a theoretical model of argument structure (Rips, in preparation). In the first structure, the argument ends when one speaker challenges the opponent's last claim, as shown in the top example in Table 1. This structure we will refer to as *Late Challenge* because the speaker presents a challenge in the last line. The second type of argument is similar to the first, except that the positions of the last line and the third line are reversed. Thus, the challenge occurs in the third line, and the argument ends with a claim. We refer to this argument type as *Early Challenge*, and an example also appears in the table. In the final argument type, the structure is again similar to that of the Late Challenge, but with the challenge omitted. As a result, the third structure

consists of only four lines, while the other two structures are five lines long. This structure will be referred to as *No Challenge* (see Table 1).

We produced these arguments by changing the order of the same group of five statements. As illustrated in Table 1, the Late Challenge and the Early Challenge arguments differed only by the interchange of the third and fifth lines. (Thus, the challenge was directed against different claims in the Late and Early arguments.) The Late Challenge and No Challenge arguments differed only by the presence or absence of the fifth line. In a few cases, it was necessary to reword individual claims in order to make the argument coherent; however, the individual meanings of the claims and queries within those structures remained constant.

Claim strength

For each of the three argument structures, we also varied which of the speakers made the weakest claim of all the claims in the argument. We altered claim strength by qualifying one of the assertions to indicate that the speaker had limited knowledge or confidence. In the examples in Table 1, Rita has the weakest claim (*I am no expert on rights or anything, but I think that abortion is not an issue of freedom*). To create arguments in which the opponent has the weakest claim, we substituted a stronger statement for Rita's weak one, and a weaker statement for one of Alan's. The following example is an argument of the Final Challenge type in which Alan presents the weakest claim (line 2):

Table 1. Sample arguments from Experiments 1 and 2

Late Challenge:

1. Rita: Abortions should not be legal.
2. Alan: But the right to have an abortion for a woman is fundamental.
3. Rita: I am no expert on rights or anything, but I think that abortion is not an issue of freedom.
4. Alan: The freedom to control your own body is one of the most important rights a person can have.
5. Rita: What is your evidence for that statement?

Early Challenge:

1. Rita: Abortions should not be legal.
2. Alan: But the right to have an abortion for a woman is fundamental.
3. Rita: What is your evidence for that statement?
4. Alan: The freedom to control your own body is one of the most important rights a person can have.
5. Rita: I am no expert on rights or anything, but I think that abortion is not an issue of freedom.

No Challenge:

1. Rita: Abortions should not be legal.
2. Alan: But the right to have an abortion for a woman is fundamental.
3. Rita: I am no expert on rights or anything, but I think that abortion is not an issue of freedom.
4. Alan: The freedom to control your own body is one of the most important rights a person can have.

1. Rita: Abortions should not be legal.
2. Alan: I am not too clear on the abortion issue, but I think that abortion might be an important right for women.
3. Rita: The right of an unborn child to live is infinitely more important than the convenience of the mother.
4. Alan: But the freedom to control your own body is one of the most important rights a person can have.
5. Rita: What is your evidence for that statement?

When the speaker presenting the challenge (Rita in the example) made the weakest claim, it was in either the last line or the third line (the third line in Late Challenge arguments and the last line in Early Challenge arguments). When the other speaker (Alan) presented the weakest claim, it was in line 2.

Initial statement

As a final variation, the arguments began either with an opening claim, as in the Table 1 examples, or with a neutral greeting that made no substantive claim. For instance, in the neutral condition, the above arguments began with Alan's statement *Hi, Rita. It's good to see you.* The rest of the argument was unchanged.

Stimulus arguments

We created the stimulus arguments by selecting 12 specific topics (e.g., chances of getting a job, college life, abortion, political issues). We then generated 12 argument versions for each topic, varying the structure, the speaker with the weakest claim, and the presence of a neutral comment, as we have just described. A subject's booklet contained one argument for each of the structure-by-strength-by-comment conditions, and the subject saw each argument phrased in terms of a distinct topic. Across subjects, however, each topic appeared equally often in the 12 conditions. We randomized the order of the arguments in the booklet for each subject.

Procedure

The first page of the booklet contained instructions and a sample argument. Subjects read the instructions on their own. The instructions told them that they would see a series of arguments and that they were to decide 'which of the two people has got more work to do in order to prove' that he or she is correct. We used the 'more-work-to-prove' wording instead of 'burden of proof' to avoid confusing subjects over the legal definitions of the latter term. Subjects circled the name of the person and then rated how confident they were in their choice by circling a number on a scale from 1 to 7 (1 = 'extremely low confidence' and 7 = 'extremely high confidence'). Analysis of the choices weighted by the confidence ratings produced the same results as did analysis of the choices alone, so we will not discuss the confidence ratings in presenting the results. A session typically took between 15 and 20 minutes.

Subjects

Forty-eight subjects participated in the experiment in order to fulfil a requirement in an introductory psychology course. None of the subjects had previously taken a course in formal logic, and all were native English speakers. We tested subjects in small groups of up to six people.

Pretest

Every argument had a single claim that we designed to be weaker than the others, as we have discussed earlier. We conducted a pretest to determine whether subjects did, in fact, perceive this claim as weakest. Subjects received 1 of 4 randomized orders of all 106 statements from the main experiment, and after each statement they rated 'how weak or strong the statement would be if it appeared in an argument'. They were asked to circle a number on a scale of 1 to 5, 1 being 'very weak' and 5 being 'very strong'. They were told that 'each statement has nothing to do with the other statements around it. Judge each one as if you had not seen the others.' There were 23 participants in the pretest, none of whom served in the main experiment. The mean rating for weak claims was 1.64, while the mean rating for other claims was 3.17, $t(22) = 8.44$, $p < 0.01$. Subjects rated the designated weak claim lower than the other claims for all topics and versions.

Results and discussion

Our main goal was to determine whether taking an initial stand (e.g., *Abortion should not be legal*) or issuing challenges (*What's your evidence?*) caused subjects to perceive a change in burden of proof. Figure 1 presents the data relevant to these questions, plotted in terms of the proportion of trials on which subjects chose the first speaker as having the heavier burden.

Effect of initial claim versus initial neutral comment

When the first speaker began the dialogue with a claim (e.g., *Abortions should not be legal*), that speaker thereby introduced the topic of the argument and took a stand on it. When the first speaker began the dialogue with a neutral comment (e.g., *It's good to see you*), it was the second speaker who took the topic-setting stand. The effect of taking the initial stand thus appears in the figure as the difference between curves with open symbols (first speaker starts with claim) and curves with filled symbols (first speaker starts with neutral comment). When the first speaker began with a claim, this speaker received the burden of proof on 55% of trials; when the first speaker began with a neutral comment, the speaker received the burden on 46% of trials. This difference is reliable $\chi^2(1) = 9.94$, $p = 0.002$, according to a repeated measures analysis for categorical data (Agresti, 1990; Koch, Landis, Freeman, Freeman and Lehnen, 1977).

Figure 1 shows, however, that the effect of taking the initial stand strongly depends on which speaker has the weak claim (e.g., *I am no expert on rights or anything, but . . .*). Consider first the case in which Speaker 1 has the weak claim (top two curves in Figure 1). In this situation, subjects' assignment of burden of proof shifts by 17.6 percentage points toward the first speaker when this speaker makes an initial claim versus a neutral comment. However, when the second speaker has the weak claim (bottom two curves in Figure 1), the difference is much smaller. Here, burden of proof shifts by only 1.4 points toward the first speaker when this speaker makes an initial claim versus comment. The interaction between the presence of a neutral comment and the strength of the claims was significant in the analysis just mentioned, $\chi^2(1) = 6.17$, $p = 0.013$. We note that in this experiment the initial claim was never itself the weak claim; the weak claim always appeared at a later, supporting position in the argument. This interaction was an unexpected one, and an interpretation is necessarily ad hoc. One way to think about the finding, however, is

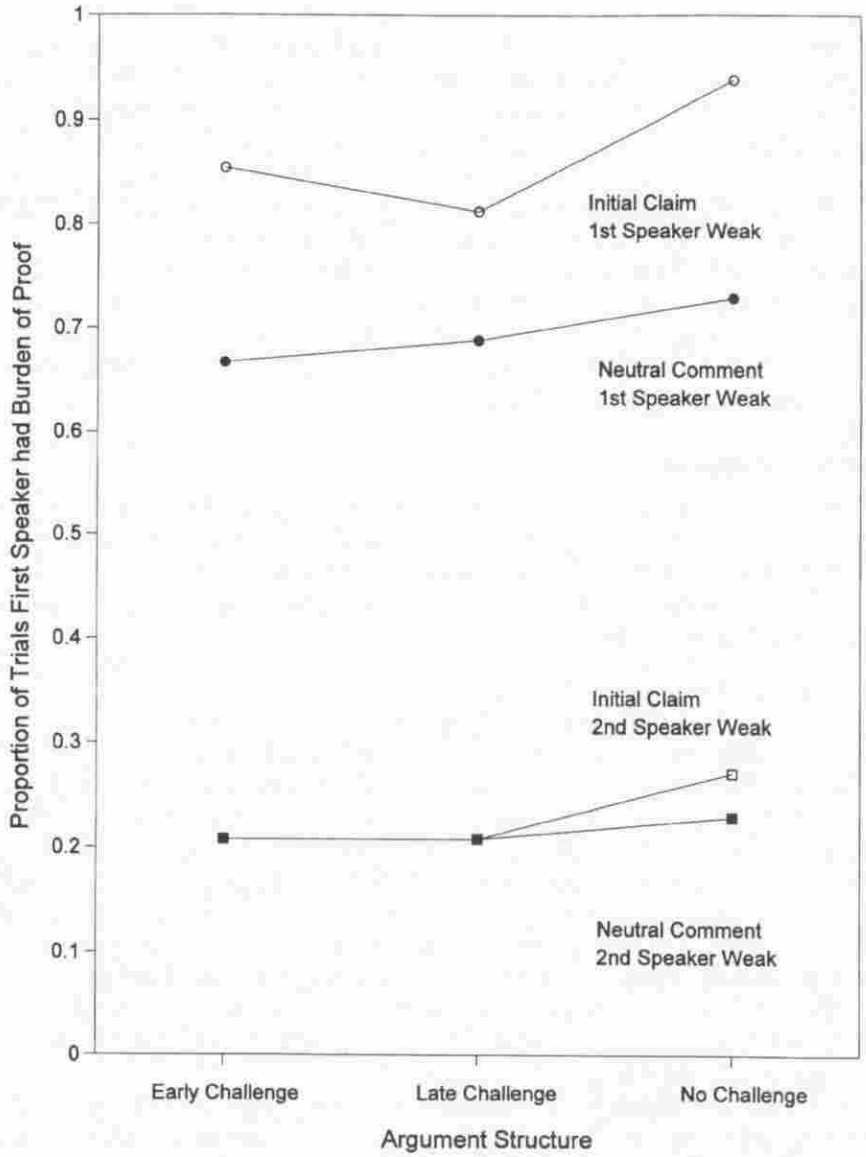

Figure 1. Proportion of trials on which subjects selected the first speaker as having the greater burden of proof, Experiment 1. Filled symbols denote arguments beginning with a neutral comment, unfilled symbols arguments beginning with a substantive claim. Circles indicate arguments in which the first speaker had the weak claim, and squares arguments in which the second speaker had the weak claim.

to imagine that the salience of an initial claim enhances its need for support. Making a claim at the very beginning of an argument strongly suggests that the speaker is capable of backing it; hence, failing to do so magnifies the speaker's burden. Claims made later in the argument are less salient and may incur less risk on the part of the speaker. If so, failing to back such a claim is not so consequential.

Not surprisingly, weak claims also produced a sizeable main effect, increasing the speaker's burden of proof. The first speaker had the burden on 78% of trials when that speaker made the weak claim, but on only 22% of trials when the second speaker made the weak claim, $\chi^2(1) = 40.14$, $p < 0.001$. However, there was no significant interaction of strength with argument structure, nor any significant triple interaction in our analysis.

Effect of challenges

In the Early and Late Challenge structures, the first speaker receives the challenge when there is a neutral comment, and the second speaker receives the challenge when there is an initial claim. If challenges increase burden of proof, we should observe an interaction between argument structure (challenge versus no challenge) and type of first line (claim or comment). Specifically, in the neutral comment condition, challenges should increase the first speaker's burden of proof; in contrast, in the initial claim condition, challenges should decrease the first speaker's burden. Figure 1, however, shows little evidence of such an interaction. When the first line is a neutral comment, subjects chose the first speaker as having the burden of proof in 44% of arguments with challenges and 48% of the arguments with no challenges. When the first line was a claim, the first speaker had the burden of proof on 52% of arguments with challenges and on 60% of arguments with no challenges. Neither the difference among argument types nor the interaction of argument and type of first line was significant in this analysis (for the main effect of argument type, $\chi^2(2) = 3.73$, $p > 0.10$; for the interaction $\chi^2(2) = 1.97$, $p > 0.10$).

These findings suggest that anti-primacy in our earlier study may have been due to the first speaker taking the issue-setting stand for the following debate. The present evidence indicates that starting the argument with a substantive claim makes the speaker susceptible to burden of proof when support for this claim is weak. In our earlier study, the evidence supporting the initial claim was probably of intermediate strength, relative to the arguments here, and this may have been sufficient to tag the first speaker as having the greater burden. The present results also show a clear-cut main effect of strength. Admitting uncertainty or lack of knowledge drastically increased the perceived burden on the speaker. The current findings also indicate that the presence of challenges probably has a less important role, since we found no significant effect of challenges in the current study. This does not mean, however, that challenges play no role at all in determining burden of proof. It is possible that subjects did not bother to process the challenge cues in the present experiment, since they could make their decision simply on the basis of the marked differences in strength of individual claims. Experiment 2 checks for effects of challenges using a more sensitive forced-choice paradigm.

EXPERIMENT 2: EFFECTS OF CHALLENGES

During this experiment, subjects choose on each trial which of three arguments gives one of the speakers the greater burden of proof. For example, a subject might see the three arguments in Table 1 and decide in which of the three arguments Rita has the greatest burden. Within each argument triple, the speaker who has the weakest claim is fixed (e.g., Rita has the weakest claim in the three arguments in Table 1). There are

also no neutral comments of the sort we used in Experiment 1. The point of this was to hold constant known factors affecting burden of proof within a triple in order to focus subjects' attention on the presence or absence of challenges. If challenges, such as *What's your evidence?*, increase the burden on the challenged speaker, we would expect subjects to select the Early Challenge or Late Challenge arguments over No Challenge when asked to pick the argument in which the speaker has the greatest burden.

Method

Subjects received a booklet consisting of five pages: an instruction page and four pages of arguments. On each page there were three versions of the same argument—Early Challenge, Late Challenge, and No Challenge—similar to the three versions in Table 1. Each argument triple was based on one of four topics that we had selected from those of the first experiment.

For each argument triple, we named one of the speakers and asked subjects to decide in which of the three arguments this speaker had 'the most to do' to prove him- or herself correct (e.g., *In which of the above arguments does Alan have the most to do in order to prove he is correct?*). We asked a subject about the first speaker on two trials and about the second speaker on the remaining two trials. Within each of these conditions, a subject saw one argument triple in which the first speaker had the weakest claim and one triple in which the second speaker had the weakest claim.

A subject saw each triple framed in terms of a different argument topic. Across subjects, however, topic was balanced with the speaker who had the weak claim and with the speaker who the subjects evaluated. We randomized the order of the four argument triples in a new way for each booklet, and we presented the items within a triple in one of two random sequences.

Twenty-one undergraduates (native English speakers) participated in the experiment. We tested subjects in groups of up to eight people, and the sessions took about 10 minutes.

Results and discussion

In the stimulus arguments, the first speaker (e.g., Rita in Table 1) always offered the challenge. If challenges are effective in shifting the burden of proof, then the first speaker's burden should be lighter in the arguments with a challenge than in the one with no challenge. Thus, subjects should pick the No Challenge argument as giving the first speaker the greater burden. The opposite pattern should apply to the second speaker: Subjects should pick one of the challenge arguments as giving the heaviest burden of proof.

Figure 2 shows qualified support for this prediction. The figure plots the proportion of times subjects selected the Early Challenge, Late Challenge, or No Challenge argument as giving greatest burden of proof to the first speaker (squares) and to the second speaker (circles). The proportions sum to one along each of the two curves. For the first speaker, subjects selected the No Challenge argument more often than the Early Challenge, and they selected the Early Challenge argument more often than the Late Challenge. Choices for the second speaker exactly reversed this ordering. Subjects chose arguments with challenges significantly more often for

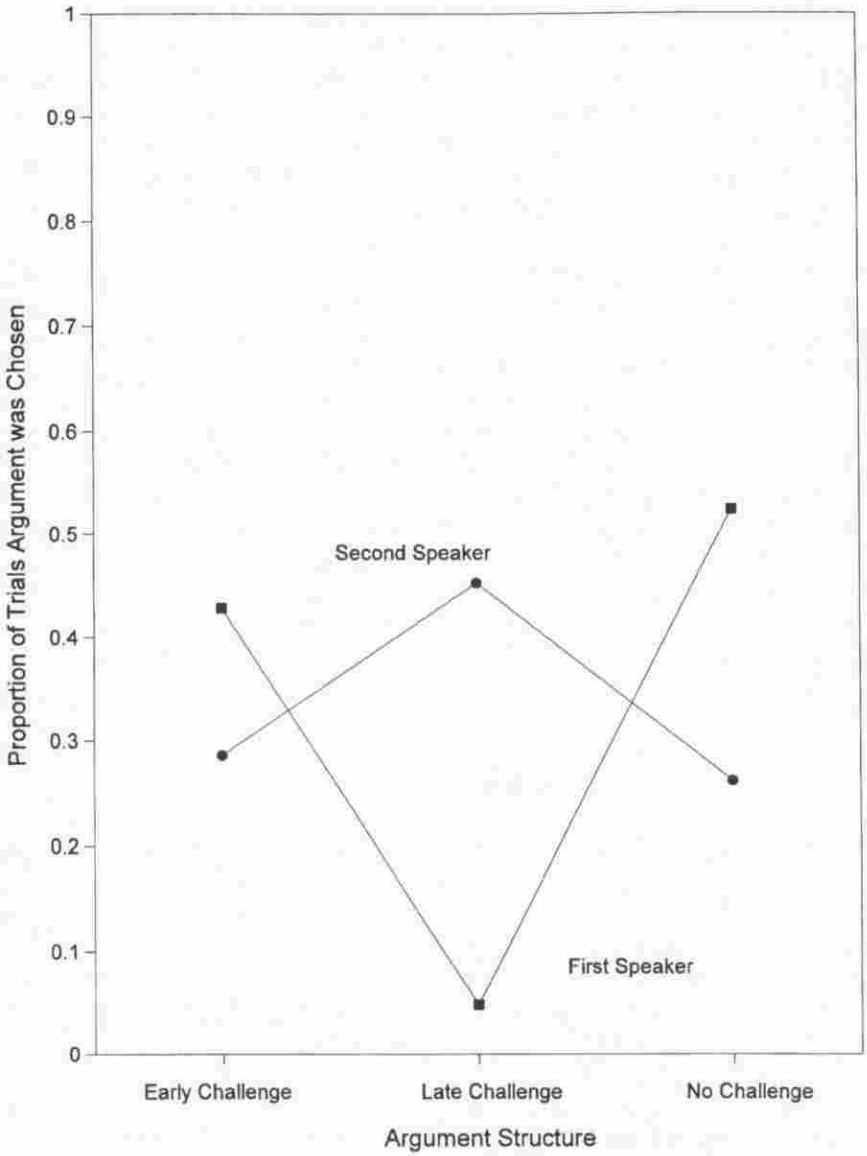

Figure 2. Proportion of trials on which subjects identified an argument (x-axis) as the one in which the first speaker (■) or the second speaker (●) had the greatest burden of proof.

the second speaker than for the first speaker ($\chi^2(1) = 4.35$, $p = 0.04$), in line with our predictions.

However, Figure 2 also shows that subjects respond differently to the two types of challenge arguments. When subjects chose which argument gave the first speaker greater burden, they selected Early Challenge over Late Challenge. Conversely, when they chose which argument gave the second speaker greater burden, they selected

Late Challenge over Early Challenge. We expected both types of challenge to increase the second speaker's burden, since it was the second speaker who received the challenge in both arguments. That is, we anticipated that Figure 2 would show the first two points on the curve for speaker 1 to be low and approximately equal and would show the first two points on the curve for speaker 2 to be high and approximately equal. Why are the Late Challenge arguments so much more effective than the Early Challenge arguments at shifting the burden to speaker 2? One possible answer stems from the fact that the second speaker has no opportunity to answer the challenge when it appears in final position. Leaving the challenge hanging in this way may have worked strongly against the second speaker.

These results suggest that subjects can pay attention to challenges when assessing burden of proof. No such effect appeared in Experiment 1, perhaps because differences in the strength of individual claims masked the challenge effect. In Experiment 2, the same speaker had the weakest claim in all three arguments within a triple. Thus, subjects could not base their choice on strength. There was no significant effect of strength across triples [$\chi^2(1) = 0.52$, $p > 0.10$], and no interaction between strength and the presence of challenges [$\chi^2(1) = 0.03$, $p > 0.10$]. Experiment 1 showed that the speaker who takes the first stand is normally at a disadvantage. The present experiment suggests that a good strategy for this speaker is to put the opponent on the defensive by challenging claims that are difficult for the other speaker to defend.

GENERAL DISCUSSION

The present studies indicate that speakers who introduce the initial claim in an argument increase their perceived burden of proof. Likewise, speakers who offer weak claims that lack conviction also appreciably increase their own burden, according to the results of Experiment 1. However, speakers can also shift the burden to their opponents in some circumstances. Experiment 2 showed that, in the absence of other cues, speakers who challenge their opponent's claims at the conclusion of an argument decrease their own burden at the expense of their opponent.

As we noted earlier, past studies in the persuasion literature have found both primacy and recency effects for sequences of evidence (Hogarth and Einhorn, 1992; McGuire, 1985; Nisbett and Ross, 1980). Primacy, in this context, means greater influence or persuasiveness of the initial evidence, and recency greater influence for the final evidence. Our own results demonstrate a greater burden on the speaker making the first claim, and we characterized this effect as anti-primacy on the grounds that this speaker is at a disadvantage during the course of the debate. We can also view the effect of a final challenge as a type of recency effect, in the sense that the speaker offering this challenge has an advantage over his or her rival, according to the results of Experiment 2. However, it would be a mistake, we believe, to view these effects as due merely to the serial order of statements in the argument. Experiment 1 demonstrated that the initial statement increases a speaker's burden of proof only when that statement is a substantive claim and only when later statements by that speaker are relatively weak. Moreover, Experiment 2 showed that the last

statement decreases a speaker's burden of proof only when it challenges the evidence on the opposite side of the issue.

These complexities arise because the arguments we have studied have a more differentiated structure than do simple sequences of pro and con evidence. During a two-person argument, speakers attempt to marshal support for earlier claims by offering further claims as reasons. They also challenge their opponent's claims or attack them outright, so the argument as a whole develops as an interlocking hierarchy of supporting and defeating relations. For example, we can view the arguments in Table 1 as hierarchical structures in which Rita's initial claim (*Abortions should not be legal*) sets the basic debating point for the argument. Alan's reply (*But the right to have an abortion for a woman is fundamental*) then becomes an attack on the preceding claim. Since this response is itself a claim, it can prompt a counterclaim (as in the Late and No Challenge arguments) or a challenge (as in the Early Challenge argument). The arguments continue in this way, obeying conversational conventions that govern the type of response appropriate at each stage of the discourse (Grice, 1989; Lewis, 1979).[2]

Viewed from this perspective, the effects we have reported may have relatively little to do with primacy and recency. Although the person introducing the first claim has a handicap, this is probably due, not to (anti-)primacy as such, but to the fact that this claim is the main issue for the entire debate. Similarly, the effect of having a challenge at the end of the argument may be due, not to recency, but to the fact that this challenge necessarily goes unanswered. These hypotheses require further refinement and testing, of course, but they seem reasonable possibilities based on our current evidence.

We do not doubt that people sometimes evaluate list-like evidence. Judgments about which applicant to accept for a job or for a graduate fellowship may be examples, since the evidence often comes in discrete pieces with little overall structure (e.g., GRE scores, GPAs, institution granting the undergraduate degree). In this context, the search for serial-position effects may serve an important purpose. We conjecture, however, that judgments based on unorganized evidence may be the exception and that most of the critical evaluations we perform we base on evidence deeply embedded in conversational structures. People may even have difficulty distinguishing evidence from other forms of information that these structures convey (Brem and Rips, 1995; Kuhn, 1991; Ranney, Schank, Hoadley and Neff, 1994). Evidence in trials, political debates, scientific disputes, negotiations in business, and discussions among family members or friends virtually never take the form of merely sequential detail. Within these disputes the verbal-learning notions of primacy and recency may be out of place. After decades of research on primacy and recency, investigators in the 1960s discovered that there were *no* standard serial position effects for words embedded in sentences (e.g., Mandler and Mandler, 1964). It may turn out, in a similar way, that there are no primacy or recency effects for evidence, except as byproducts of their embedding arguments. If that is true, then models of belief revision based on simple linear adjustment and anchoring may account for at most a modest sample of judgments and greater effort should be spent examining evidence in its natural habitats.

[2]Rips (in preparation) is a theory of this argument structure.

REFERENCES

Agresti, A. (1990). *Categorical data analysis*. New York: Wiley.

Anderson, N. H. (1981). *Foundations of information integration theory*. London: Academic Press.

Brem, S. and Rips, L. J. (1995). Explanation and evidence in informal reasoning. *Proceedings of the 17th Annual Conference of the Cognitive Science Society*, 271–276.

Chaiken, S. (1979). Communicator physical attractiveness and persuasion. *Journal of Personality and Social Psychology*, **37**, 1387–1397.

Clark, H. H. and Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science*, **13**, 259–194.

Gaskins, R. H. (1992). *Burdens of proof in modern discourse*. New Haven, CT: Yale University Press.

Grice, H. P. (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.

Hogarth, R. M. and Einhorn, H. J. (1992). Order effects in belief updating: the belief-adjustment model. *Cognitive Psychology*, **24**, 1–55.

Hovland, C. I. and Weiss, W. (1951). The influence of source-credibility on communication effectiveness. *Public Opinion Quarterly*, **15**, 635–650.

Kelman, H. C. and Hovland, C. I. (1953) "Reinstatement" of the communicator in delayed measurement of opinion change. *Journal of Abnormal and Social Psychology*, **48**, 327–335.

Kieras, D. E. (1978). Good and bad structure in simple paragraphs: effects on apparent theme, reading time, and recall. *Journal of Verbal Learning and Verbal Behavior*, **17**, 13–28.

Koch, G. G., Landis, J. R., Freeman, J. L., Freeman, D. H. and Lehnen, R. G. (1977). A general methodology for the analysis of experiments with repeated measurement of categorical data. *Biometrics*, **33**, 133–158.

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

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

Mandler, G. and Mandler, J. M. (1964). Serial position effects in sentences. *Journal of Verbal Learning and Verbal Behavior*, **3**, 195–202.

McGuire, W. J. (1985). Attitudes and attitude change. In G. Lindzey and E. Aronson (Eds.), *Handbook of social psychology* (vol. 2, pp. 233–346). New York: Random House.

Nisbett, R. and Ross, L. (1980). *Human inference: strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice-Hall.

Ranney, M., Schank, P., Hoadley, C. and Neff, J. (1994). "I know one when I see one". How (much) do hypotheses differ from evidence? *Proceedings of the 5th ASIS SIG/CR Classification Research Workshop*, 139–155.

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

Rips, L. J. (in preparation). Reasoning and conversation.

Schiffrin, D. (1987). *Discourse markers*. Cambridge: Cambridge University Press.

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

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