UC Merced

Title
Claim Strength and Burden of Proof in Interactive Arguments

Permalink
https://escholarship.org/uc/item/5jp8023q

Journal

Author
Bailenson, Jeremy N.

Publication Date
1997

Peer reviewed

Claim Strength and Burden of Proof in Interactive Arguments

Jeremy N. Bailenson
Department of Psychology
Northwestern University
Evanston, IL, 60640
jnb@nwu.edu

Abstract

Previous research shows an *anti-primacy* effect (Bailenson & Rips, 1996), in that the first speaker in a conversational argument incurs more Burden of Proof (BOP) than the second speaker. In addition, claims may be encoded differently when they are embedded in a structured dialogue than when processed outside the context of the argument. We were interested in determining how the strength of specific claims in the argument depend on their location in the structure as a whole, and whether anti-primacy would persist in disputes where the claims offered by the two speakers were equally convincing. Subjects read interactive arguments between two speakers having a conversation, rated the convincingness and support levels of the individual claims when they were both embedded in the dialogue and removed from the dialogue, and judged overall burden of proof. Different groups of subjects saw the same arguments, the only difference being which speaker (first or second) made the particular groups of claims. The anti-primacy effect occurred even though the strength of the claims did not change as a function of which speaker presented them. In addition there was no difference between convincingness and support ratings, although the results demonstrated that the level of both types of ratings was somewhat a function of where in the argument structure the claims were situated. Specifically subjects perceived claims occurring in the initial position in the dialogue as less convincing than the same claims when the claims were removed from the context of the argument. Furthermore, these initial claims correlated less with BOP than did the later claims.

Introduction

Participants of a conversational dispute, as well as the observers, make judgments as to which of the opposing sides is doing a better job arguing their position. When people in an argument make claims, they must defend those claims with accepted evidence. After making a series of statements, each speaker has some degree of *burden of proof*. In other words, the participants have to do a certain amount of persuading in order to prove they are correct. If one speaker presents claims that are far more convincing than the other speaker, then that individual would have little to do in

order to prove his or her position. The burden would lie on the other participant, who would then be forced to overcome the presumption created by the disparity between the quality of the two positions. In this sense, the BOP shifts from one speaker to the other as the argument progresses, depending on certain rules (Rips, 1996; Gaskins, 1992).

One of the factors BOP is based on is the convincingness levels of the individual claims in an argument. Previous researchers have defined claim strength in many different ways ranging from size of the claim (Gulley & Berlo, 1956) to how "bold" it is (Baron, 1991). Strength in argument can be approached in a number of ways, but in this paper we would like to draw the distinction between *local* and *global* strength. Local strength concerns the inherent characteristics of individual claims which make them convincing. Conceivably, the local strength of a claim is the same regardless of its context--if isolated from the other surrounding claims in the argument then the local strength should not change. 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 the previous claim, and how well it supports the speakers position as a whole. Intuitively, Burden of Proof should depend on both of these measures of strength.

While people who judge disputes clearly pay attention to strength of the claims in an argument, other research shows that they are affected by structural factors that are independent of the meaning of the claims. Presenting two groups of identical claims in different orders can produce two separate patterns of judgment (McGuire, 1957). For example, some studies find a recency effect, in that information at the end of an argument is retained better in memory and is more persuasive than information at the beginning of an argument. These effects have been found in descriptions of U.S. presidents (Anderson, 1973), 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 information from the beginning of an argument more effectively, and this information then biases how later claims are processed. Pennington and Hastie showed this phenomenon in mock jurors: The first bit of information the jurors received initiated their mental story representations of the supposed crime, thus framing how the proceeding

claims were interpreted such that they were consistent with the overall story (1986). Primacy effects also occur in attributions of how other people perform a task (Benassi, 1982) and for judgments of contingencies (Yates & Curley, 1986).

In a series of experiments Bailenson and Rips (1996) found an "anti-primacy" effect, in that the speaker who initiates the debate incurs more burden of proof then the second speaker in arguments where speakers present the same number of claims. This effect occurs when the first speaker makes the final claim in the argument and also when the second speaker makes the final claim. This first speaker bias seems to be an effect of position which only surfaces in a structured dialogue, as previous studies, while demonstrating primacy and recency (Hogarth & Einhorn, 1992), have not shown a disadvantage in persuasion for items at the top of a serial list. One plausible explanation for this effect may have to do with the roles of the two speakers. In work designed to study how people plan conversations, Scholtens (1991) describes the difference between the initiator of the conversation and the recipient. The initiator (i.e. speaker 1) "stages" the argument, in that he or she provides the underlying structure or script that will be followed for the remainder of the argument by the recipient (i.e. speaker 2). The recipient does not have much control in molding the hierarchy of the conversation, as his or her job is primarily geared towards following the initiator's argument: collecting information, signaling departing ideas, pointing out obscurities and falsities, and giving general reactions.

In the current study, we attempted to test whether or not we could demonstrate the anti-primacy effect in arguments where the claims made by each speaker were equally convincing. We wanted to ensure that the first speaker's claims did not become less convincing than those of the second speaker just because they were made by initiator of the argument. In this experiment subjects assessed disputes in which we manipulated the serial position of each side of an argument, in that the same claims could be shifted to accommodate either the role of first speaker or second speaker. In other words, if in one argument the first speaker argues in favor of abortion while the second speaker was argues against it, then we designed another argument in which the claims were identical to the ones used in the other version except that the first speaker now argues against abortion. It would be impossible to argue that anti-primacy results from the type of claims offered by two different speakers if all of the claims appear in both positions. In addition we gauged the strength of particular claims when they were situated within the argument structure and when they were removed from the structure. We predicted that anti-primacy would persist even though the claims are identical across subjects for the two speakers, as the first speaker's role as initiator sets the forthcoming context for the argument independently of the convincingness of the claims. We also predicted that strength would vary as a function of the location of the claim in the argument structure, due to the unique characteristics of arguments in dialogue.

Method

Materials

We constructed booklets of nine different arguments consisting of six statements apiece. Each argument appeared on a separate page. The content of the arguments varied, but typically involved entertainment, politics, or college life as the following example illustrates:

> 1. Pat: Baseball has more breaks in the action than other sports.
> 2. Jim: Baseball has fewer breaks in the action than other sports.
> 3. Pat: You have to sit through all of the side changes.
> 4. Jim: The only long break comes after the seventh inning.
> 5. Pat: Every time they substitute pitchers there is another fifteen minute break.
> 6 Jim: At least there are not many substitutions in baseball compared to other sports.

Each argument appeared in two possible orders. We produced the alternative order of each argument by reversing the positions of claims 1 and 2 (*initial claims*), claims 3 and 4 (*middle claims*), and claims 5 and 6 (*final claims*). To produce the alternative version of the argument above, we changed the orders of the claims as follows:

> 1. Jim: Baseball has fewer breaks in the action than other sports.
> 2. Pat: Baseball has more breaks in the action than other sports.
> 3. Jim: The only long break comes after the seventh inning.
> 4. Pat: You have to sit through all of the side changes.
> 5. Jim: At least there are not many substitutions in baseball compared to other sports.
> 6. Pat: Every time they substitute pitchers there is another fifteen minute break.

As a result, there were 18 different arguments, 2 different versions for each of the 9 different topics.

We wanted to ensure that the claims made by each speaker in the argument were roughly equivalent in strength. In other words, we wanted the total strength of the three claims made by Jim in the above argument to equal the total strength of the three claims made by Pat. So after creating the arguments, we then ran a pretest in order to determine the strength of the claims made by each speaker.

Design

Each participant received two booklets, each one with all of the nine argument topics. In one booklet, we asked them to choose which speaker had more burden of proof, and in the other they we asked them to rate the strength of each

14

claim. The order of the booklets was counterbalanced, so that half of the subjects rated the strength of the claims first and half of the subjects chose which speaker had more burden of proof first. Each of the two booklets contained exactly the same arguments (same versions of all of the 9 topics) in the same order. So each subject rated claim strength and burden of proof on the same nine arguments. In this manner we could compare the same subjects' ratings of strength and burden of proof or we could only analyze the data from the first booklet in order to test the effects the two measures had on each other.

We utilized two different methods of determining strength in the arguments, one which gauged the convincingness of each claim (*local strength*), and one which determined how well the claims fit into the argument as a whole (*global strength*). In the booklet in which subjects were asked to rate the strength of the claims, we used two different questions. Half the subjects rated "How convincing is the claim, in and of itself?" The remaining subjects rated "How well does the claim fit into the argument as a whole?" Each subject saw the same type of strength question for all nine arguments in the strength rating booklet.

For each topic subjects saw only one of the versions (i.e. subjects never saw the same argument topic twice with different speakers making the initial claim). Every subject saw the topics in a different random order. Across subjects, each version for each topic appeared an equal number of times.

Procedure

The first page of the Burden of Proof booklet contained instructions and a sample argument. Subjects read the instructions on their own. We 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 they are correct." They were instructed to circle the name of the person who needs to do more to prove that he or she is correct. Also, we asked them to rate how confident they were in their choice, by circling a number on a scale from 1 to 7, 1 being extremely low confidence and 7 being extremely high confidence. The layout of the page was as follows: the argument appeared, followed by the names of the two speakers underneath it and the rating scale. Each argument appeared on the same page

The first page of the local strength booklet instructed subjects to read each claim, and to decide how convincing it is. We instructed them to ignore how well the claim fit into the argument as a whole. They rated each of the six claims in the argument on a 7 point scale, with 1 corresponding to an unconvincing claim, and a 7 corresponding to a very convincing claim.

The first page of the global strength condition instructed subjects to read each claim, and to evaluate how well the claims support the speakers argument by relating it to the other claims. We asked them "How strongly does each claim support the speaker's position?" If a claim fit well in the argument as a whole then it gave strong support. They rated each of the six claims on a 7 point scale, with 1 corresponding to a claim which offered weak support to the argument, and a 7 corresponding to a claim which offered strong support to the argument as a whole.

Subjects all received the BOP booklet and one of the two strength rating booklets (not necessarily in that order). No subject began the second booklet until everyone in the session was finished with the first one. Sessions typically took about twenty minutes.

Participants

48 subjects participated in the experiment in order to fulfill a requirement in an introductory psychology course. 24 subjects received the local strength booklet, and 24 subjects received the global strength booklet. All subjects were native speakers of English. Subjects were tested in small groups of up to 10 people.

Pretest

We ran a pretest for two reasons. First, we wanted to ensure that for all of the argument topics, one speaker did not present claims which were vastly more convincing than the other, since we did not want any potential effects to be overwhelmed by extreme differentials in the strength of claims in the opposing positions. Second, we wanted to have subjects rate all the claims in isolation so that we could provide a baseline with which we could compare the claims when they were embedded in the actual argument structure.

After constructing the initial arguments, we took the claims out of the argument structures and randomized them such that claims from the 9 different topics were mixed. In addition, we inserted "fillers", or claims which were slight variations of the original claims, that made them more or less convincing. These fillers could be substituted for the original claims in order to adjust the strength of a given speaker's total position.

The instructions to the pretest were similar to the local strength condition in the main experiment. We asked subjects to judge how convincing each claim was on a 7 point scale. The claims were presented in four different random orders. Participants were introductory psychology students, and sessions took about twenty minutes.

We averaged the ratings of strength of the claims offered by each speaker (3 claims for 2 speakers in each argument). In six of the nine arguments it was necessary to use filler claims in order to equate the strength of the positions offered by the two speakers. After these substitutions were made, the average difference in position was .16 and the median difference was .15 on a scale from 1 to 7. The largest difference was .32, $t(31)=1.45$, $p>.05$. The pretest ensured that any differences in BOP could not be attributable to differences in the strength of the speaker's position as a whole.

Results

Across conditions, subjects chose the first speaker as having more burden of proof (anti-primacy) in 58% of the arguments. This difference was reliably different from chance, $t(47)=3.28$, $p<.05$. However, as Table 1 demonstrates, convincingness ratings for any given claim

was the same regardless of whether it was made by speaker 1 or by speaker 2, $F(1, 47)=.02$, n.s. It is important to note here that in the order manipulation we altered which speaker (first or second) made a particular claim. We did not shift the overall position of any claims. So, if in one order condition an

Table 1: Mean strength ratings by position for Speaker1 and Speaker 2. Ratings are averaged across Global and Local conditions.

Position	Speaker	
	1	2
Initial Claims	2.67	2.58
Middle Claims	4.49	4.47
Final Claims	4.73	4.83

initial claim is made by speaker 1, in the other version that claim is made by speaker 2, but it still remained in the initial position. In other words, the position (initial, middle, final) of the claims never changed. Strength ratings varied according to the position of claims, $F(2, 47)=88.77$, $p<.05$. Initial claims were rated as weakest (2.67), followed by middle claims (4.47). Subjects rated the final claims the highest (4.76). There was no effect for type of strength rating, $F(1, 47)=.02$, n.s. Subjects rated a claim equally whether they were asked to rate convincingness in and of itself (local = 3.98) or to rate how well it supported the argument as a whole (global = 3.95). However this null effect may be a result of subjects not fully grasping the instructions describing the difference between the two types of ratings.

We ran a regression between strength ratings and BOP scores in order to test which parts of the argument were more predictive of BOP choices. These correlations are all within subject, in that the same subjects are rating both strength and judging BOP. The correlation between strength of initial claims and BOP was not significant, $r=.14$. The correlations between middle claims and final claims were both significant, $r=.32$ and $r=.39$ respectively, $p<.05$. Initial claims in an argument, in addition to being less convincing than the other claims, are less predictive of burden of proof.

It could be the case that claims in the middle and final position are rated stronger then claims in the initial position because of the content of the claims or because of the positioning of the claim in the argument. To resolve this issue, we used the pretest data to factor out the strength of the claims when they were rated in a random order from the strength ratings when they were rated in the actual argument structure (we averaged across global and local ratings). Using a regression with claim number as a variable (1 through 6), we found that claims later in the argument were rated stronger than early claims, even when isolated claim strength had been factored out, $t(47)=4.59$, $p<.05$. Figure 1 plots rating type (isolated or structured) as a function of where in the argument the claims occurred (initial, middle, or final). A by-item analysis demonstrates that the mean for
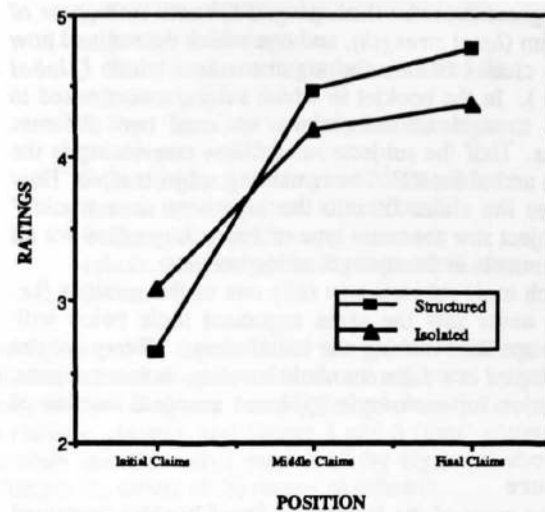

Figure 1: Comparison of Structured (main experiment) and Isolated (pretest) strength ratings.

the two initial claims in the pretest was 3.09 while in the main experiment it was 2.66, $t(8)=3.25$, $p<.05$. Initial claims were rated weaker when they were in a structured argument then when they were rated in a random order. Claims in the middle position and claims in the final position were rated as stronger in the structured arguments then in the pretest, with differences of .28 and .39 respectively, but neither of these differences were significant, $t(8)=1.56$, and $t(8)=1.32$. While it could be the case that these results are due in part to subjects having difficulty in interpreting the sentences in a mixed, random order (versus in context of the argument as a whole) in the pretest, none of the subjects expressed having any difficulty performing the task. The results demonstrate that initial claims deflate while they are in the argument structure, while the later claims slightly inflate in convincingness. So subjects perceive claims as being more convincing as the argument becomes further developed, or at the least they do not find claims which identify their role as an initiator to be as convincing as the other claims in the dispute.

General Discussion

The results show that the first speaker typically is assigned more burden of proof than the second speaker, even though

16

the claims of the two speakers are equally convincing. Subjects were not sensitive to the differences in how convincing a claim is (local strength) and how well it supports the argument as a whole (global strength) when they are rating the claims in the argument structure. In addition, the location of a particular claim in an argument affects subjects' perception of its convincingness and support levels. Claims made early in the argument are rated lower than when they are rated in isolation and the reverse holds for claims made towards the end of an argument, almost as if the strength is transferred. In addition, claims located farther down in the structure tend to correlate more with BOP than do the early ones.

It could be the case that as claims proceed in a linear order, the strength of each claim for a given speaker is passed down to that speaker's next claim such that the later claim inherits some of the strength of the former. Claims late in an argument depend on the earlier ones, in the sense that they operate on assumptions that the initial ones pioneered. Since these early assumptions are implicitly contained in the later ones, it could be the case that when assessing the strength of the late claims subjects include the contributing parts of the early ones, and as a result inflate their strength. This notion is somewhat supported by work done by Orsolini (1993), where she argues that arguments in conversation are governed by a unique sequential organization, in which an initial claim primes the other speaker in the argument and allows him or her to further process proceeding justifications. In other words, understanding the later claims depends on understanding the earlier ones. These types of linear dependency relations between claims may explain why later propositions are gauged as more convincing and correlate more with burden of proof.

Previous studies of arguments for the most part merely interchanged the positions of claims or evidence in unorganized serial lists. This type of argument, discrete pieces with little overall binding structure, finds its home predominantly in the laboratory. When evaluating arguments in discourse, as discussed above, the claims cannot be entirely dissociated from the overall configuration of the dialogue. As a result, order effects found in a list of unbound claims may not hold when those claims are embedded in a conversation. 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 some degree of BOP. Likewise, claims which would normally persuade people when removed from the context of a conversational argument may not be as convincing when appearing in the initial position of the dispute. These findings call into question studies on propositions that ignore the argument structure containing those propositions, and indicate that judging within a conversation may be unique.

These findings have implications for psychology, law, advertising, or for any fields which typically study arguments situated in dialogue. Moreover, systems which attempt artificial processing of arguments may benefit by assimilating the strategies of human judges. In future experiments we plan on testing these ideas with various

types of argument structures, and to isolate the contexts where anti-primacy and the differential ratings of claims persist.

References

Anderson, N. H. (1973). Information integration theory applied to attitudes about U.S. presidents. *Journal of Educational Psychology*, 64, 1-8.

Bailenson, J. & Rips, L. J.(1996) Informal Reasoning and Burden of Proof, *Applied Cognitive Psychology*. 10, pg. 3-16.

Baron, J. (1991). Beliefs about thinking. In J. F. Voss, D.N. Perkins, & J. W. Segal (Eds.), *Informal reasoning and education* 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. *The Journal of General Psychology*, 113, 351-357.

Gaskins, R. H., (1992). *Burdens of Proof in Modern Discourse.* New Haven, CT: Yale University Press.

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

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

Luchins, A. S., & Luchins, E. H. (1984). Conceptions of personality and order effects in forming impressions. *The 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* . New
Haven, CT: Yale Univ. Press.

Orsolini, M. (1993). "Dwarfs do not shoot": An analysis of
children's justifications, *Cognition and Instruction*,
11(3&4), pg. 281-297).

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

Rips, L.J. (1996). Reasoning and Conversation,
Unpublished manuscript.

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

Yates, J. F., & Curley, S. P. (1986). Contingency
judgment: Primacy effects and attention decrement. *Acta
Psychologica*, 62, 293-302.