

The Motivated Processing of Political Arguments

Charles S. Taber
Department of Political Science
Stony Brook University

Damon Cann
Department of Political Science
Utah State University

Simona Kucsova
Department of Political Science
Grand Valley State University

Abstract: We report the results of an experiment designed to replicate and extend recent findings on motivated political reasoning. In particular, we are interested in disconfirmation biases – the tendency to counter-argue or discount information with which one disagrees – in the processing of political arguments on policy issues. Our experiment examines 8 issues, including some of local relevance and some of national relevance, and manipulates the presentation format of the policy arguments. We find strong support for our basic disconfirmation hypothesis: people seem unable to ignore their prior beliefs when processing arguments or evidence. We also find that this bias is moderated by political sophistication and strength of prior attitude. We do not find, however, that argument type matters, suggesting that motivated biases are quite robust to changes in argument format. Finally, we find strong support for the polarization of attitudes as a consequence of biased processing.

Acknowledgements: Research funded by NSF Grant SES-931351. With special thanks to the students in the POL 608 political psychology seminar at Stony Brook University, Fall 2000.

Research on the motivated processing of political arguments has shown that we should not expect citizens to approach political arguments or evidence evenhandedly (Lebo and Cassino 2007; Lodge and Taber 2000; Lord, Ross, and Lepper 1979; Meffert et al. 2006; Redlawsk 2002; Rudolph 2006; Taber 2003; Taber and Lodge 2006; Taber, Lodge, and Glathar 2001). Conservatives watch Fox News and avoid CNN. Gun control supporters actively counterargue the rhetoric of the NRA. Evidence of failure in Iraq is accepted unquestioningly by critics of the war but summarily rejected by supporters. In short, citizens' prior attitudes toward the people, groups, or issues implicated in political arguments strongly bias how they process those arguments, through selective exposure or selective judgment processes. Such motivated reasoning, we believe, is hard to square with normative standards for belief updating, which require some independence of new evidence from priors (Evans and Over 1996; Fischhoff and Beyth-Marom 1983).

Bayes' Rule provides the best known such standard in the form of a mathematical rule for how existing beliefs should change in the light of new evidence: $P(A|B) = P(B|A)P(A)/P(B)$, where A is one's belief and B is new evidence. In essence, Bayes' Rule sets our posterior belief after hearing new evidence, $P(A|B)$, to be equal to the product of our prior belief, $P(A)$, and the likelihood that B would occur if A is true.

On the other hand, motivated biases may not violate Bayes' Rule for belief updating, which makes no stipulation at all about the subjective treatment of new evidence, defining rather how subjectively perceived evidence should be combined with priors to form posterior beliefs (Gerber and Green 1999). Perhaps it is perfectly acceptable for people to dismiss evidence that rings hollow to them. Perhaps beliefs are worth protecting against inconvenient evidence (Abelson and Prentice 1989).

We are not convinced, however, by this argument because it fails to recognize that Bayes' Rule alone does not provide a complete normative standard. Without making additional assumptions about how new evidence is to be treated, one cannot know whether Bayes' Rule will define acceptable belief updating. Leaving the treatment of new evidence fully subjective seems reasonable until one realizes that very dysfunctional behavior can then arise that will be fully consistent with Bayes' Rule (e.g., beliefs in a flat world that persist after one doesn't fall off the edge, or beliefs that smoking poses no personal risk that survive lung cancer). If there are no constraints on the acceptable treatment of new evidence, then not only does Bayes not provide a normative standard, it becomes tautological. It can account for any change in beliefs we might observe. If people can be completely unresponsive or perverse in their perceptions of new evidence, then they cannot behave rationally in response to their environments. And if societies polarize in response to a common stream of information, then it is hard to imagine the marketplace of ideas operating as an efficient arbiter of policy differences. Unfortunately, these are the normative dilemmas that motivated reasoning research presents.

This paper replicates and extends recent empirical findings on *disconfirmation biases*, processes that people can use to protect or even bolster their prior attitudes and beliefs in the face of discrepant evidence (Ditto and Lopez 1992; Edwards and Smith 1996; Kunda 1990; Taber and Lodge 2006). Do we welcome congenial information with little or no scrutiny while subjecting challenging information to the jaundiced eye of skepticism (Ditto et al. 1998)? Are we able to separate our evaluation of arguments and evidence from whether we agree or disagree with their implications for policies or candidates we care about?

A variety of findings from social and political psychology suggest not. For example, perceptions of media coverage of the Bosnian conflict were strongly affected by prior opinions about that conflict, with partisans on all sides discounting information supplied by the media as the biased product of a “hostile media” (Matheson and Durson 2001). Evaluations of arguments or evidence on capital punishment (Edwards and Smith 1996; Lord, Ross, and Lepper 1979), affirmative action, and gun control (Taber and Lodge 2006) are all strongly influenced by citizens’ prior beliefs on those issues. When motivated to discount arguments that challenge their priors, people may find a variety of ways to denigrate the source, the evidence presented, the validity of the “research design” (e.g., questioning sampling procedures in opinion data), and the logic or format of presentation. By contrast, arguments consistent with priors are likely to be accepted without much scrutiny, even when potential defects are pointed out or when counterarguments are supplied (Taber and Lodge 2006). Such findings raise serious questions about the descriptive accuracy of normative models of belief updating that assume people are fully responsive to new information.

This paper makes several contributions to the literature on motivated political reasoning. First, we replicate using a broader range of political issues and arguments several key findings: (1) an *attitude congruency bias*, where people evaluate arguments and evidence that support their priors as stronger and more compelling than contrary arguments (Redlawsk 2002; Rudolph 2006; Taber and Lodge 2006); (2) a *disconfirmation bias*, where people spend time and cognitive resources denigrating and counter-arguing attitudinally incongruent arguments (Ditto and Lopez 1992; Edwards and Smith 1996; Taber and Lodge 2006); and (3) *attitude polarization*, where these biases in

information processing promote more extreme attitudes so that citizens on opposite sides of an issue will tend to diverge after processing the same stream of information (Taber and Lodge 2006). The first of these findings is well-established, but the underlying processes of disconfirmation are not. Most important, attitude polarization, which is the most normatively troubling prediction of motivated reasoning research, has only very rarely been demonstrated convincingly. In short, disconfirmation and polarization require replication.

Second, we explore the impact of political knowledge (sophistication), attitude strength, and argument style on motivated bias as a way of exploring the limiting conditions for motivated reasoning. The cognitive and informational requirements of motivated reasoning are quite high, so we should expect sophisticated or knowledgeable participants to be particularly susceptible. Moreover, motivated biases should also vary directly with the strength of prior attitudes, that is, with the motivation to defend the prior attitude, which in our theory is a function of affect. Those with sufficient knowledge to construct counter-arguments (opportunity) and those with strong prior attitudes on the given issue (motive) should be most biased in their processing of political arguments.

In our theory (Lodge and Taber 2000; Taber and Lodge 2006; Taber, Lodge, and Glathar 2001), motivated biases are triggered by affective reactions to the evidence or arguments under consideration. Our own stronger findings of motivated bias in political information processing may be due to the use of longer, more complex, and more evocative arguments as experimental stimuli than is customary in the psychological literature. To test this possibility, we manipulate the length and complexity of arguments, with experimental participants rating short, pithy arguments like those used in typical

social psychology experiments (e.g., Edwards and Smith 1996); longer, more complex arguments like those we have used (Taber and Lodge 2006); and longer, two-sided arguments that embed a weak counter-argument. We included the two-sided argument type because there has been empirical support in social psychology for a “credibility effect,” such that proponents of an issue are more persuasive if they display awareness, understanding, and even sympathy with opposing arguments. In particular, information sources that present arguments counter to their known position or interests (as when an African-American argues against affirmative action) are generally viewed as more credible (for a review, see Petty and Wegener 1998; for a related theoretical argument, see Chong and Druckman 2007). This credibility, we reasoned, should enhance the perceived strength of an argument that provided and then refuted a weak counter-argument. These two-sided arguments should be more persuasive, and therefore should reduce motivated biases relative to one-sided arguments of equal length.

Methods and Research Design

General Design and Hypotheses

The experiment was designed in six parts, with the order of parts two, three, and four (the within-subjects experimental tasks) randomized (Figure 1):

[Figure 1 about here]

First, participants answered a battery of questions designed to measure the strength and position of their attitudes on eight issues: legalization of marijuana, gays in the military, animal testing by drug companies, the electoral college, U.S. foreign aid, university funding of unpopular groups, punishing cheaters in college, and tuition raises.

These issues provide variance in terms of scope (national versus local campus relevance), attitude strength, and attitude position. This is a larger and more diverse set of issues than have been used in previous motivated reasoning studies.

Next, participants completed three argument strength rating tasks in random order, using three issues drawn randomly from the eight issues listed above.¹ For one of these tasks, they rated short one-sided arguments; for another they rated long one-sided arguments; and for the third they rated long two-sided arguments.

We then asked a set of demographic and general political knowledge questions, followed by a second administration of the same battery of attitude questions used in the pre-test.

Our main predictions are:

- *H1: an attitude congruence bias*, such that counter-attitudinal arguments will be judged to be weaker than pro-attitudinal arguments.
- *H2: a disconfirmation bias*, such that counter-attitudinal arguments will take longer to rate than pro-attitudinal arguments and will engender more thoughts overall and more disconfirming thoughts than pro-attitudinal arguments.
- *H3: a polarization effect*, such that these processing biases will cause attitudes to become more extreme after rating a balanced stream of pro and con arguments.

We expect that all three of these processes will be moderated by the strength of attitudes and the sophistication of respondents, so that those with stronger attitudes and

¹ Ideally, each subject would rate arguments on every issue (8) in every argument style (3), but such a fully factorial design would have taken longer and been more demanding on subjects (entailing 128 arguments to be rated!) than we were comfortable with, so we randomly selected three issues, one for each argument style, for each subject.

greater political knowledge will be more prone to motivated bias. Moreover, we predict that higher levels of bias in processing arguments will drive greater polarization.

We also expect that these processes will be contingent on the style of presentation of the arguments.

- *H4a: an argument complexity effect*, such that longer, more complex one-sided arguments will generate more motivated bias than shorter, simpler arguments.
- *H4b: an argument credibility effect*, such that two-sided arguments, which embed a weak counterargument, will generate less motivated bias than one-sided arguments of equal complexity.

Participants

One hundred and sixty-four undergraduate students, recruited from political science classes at Stony Brook University, participated in this study for course extra credit or to fulfill subject pool requirements for political science majors. The sample was young (95% < 22), 51% female, 58% non-white, 65% moderately interested in politics, and leaning Democratic (47% Democrat: 20% Republican; 33% independent or don't know).

Stimuli, Procedures, and Measures

The experiment was conducted on personal computers in individual experimental rooms, with the pace of the experiment controlled by the participant. All instructions and tasks, with the exception of thoughts listing, were computerized. Arguments were modified from the policy debates found on internet sites or published media sources.²

² Though we did roughly match the length and complexity of arguments within a type, there is some variance on both. Argument length can be positively related to strength ratings, and indeed in our data

Table 1 presents two example arguments in each style. Note that the basic premise of each argument is identical; long one-sided arguments elaborate on the point made in the first sentence, which is always the same as the short argument; two-sided arguments replace the middle sentence of long arguments with a counter. All arguments may be found in the online Appendix.

[Table 1 about here]

In the attitude battery administered before and after the argument rating tasks, participants were asked where they stand on each of the eight issues (question wording is reported in the online Appendix). Their responses, which they indicated by sliding a scroll bar anchored by the labels “Oppose” and “Support”, provide a 101 pt thermometer measure of their pre and post attitudes on these issues. These were immediately followed by four agree/disagree items for each issue (each with a 9 pt Likert type response set), which we combined into an additive scale, our second measure of position of prior and posterior attitudes.³ Four standard measures of attitude strength (“How much do you personally care,” “How strong are your feelings,” “How certain are your feelings,” and “How much thinking have you done?”) followed for each issue (Petty and Krosnick

length weakly predicts rated strength (in bivariate regressions of strength ratings on word counts conducted separately for the three argument types). However, overall length is balanced between pro and con arguments, so this marginal bias does not affect our results. Keep also in mind that length varies by definition between argument types, where short arguments average 15-20 words and long and two-sided arguments average 30-40 words. We present twice as many short arguments to balance the overall amount of information rated for each argument type.

³ With the exception of university funding of unpopular groups, these all are clearly reliable scales (α 's > .7). For that one issue, the reliability coefficient dipped to .5.

1995). We combined these into an additive scale for attitude strength on each issue.⁴ As in prior research (Petty and Krosnick 1995; Taber and Lodge 2006), attitude strength and (folded) attitude position correlate only weakly and form largely independent attitudinal dimensions.⁵

After completing the pre-test attitude battery, subjects entered the experimental phase in which they rated the strength of arguments on three of the eight issues (randomly chosen), each presented in one of the three argument types (in random order). To hold relatively constant the amount of information in each condition, we presented 8 short arguments, 4 long arguments, and 4 two-sided arguments. For example, one participant might rate four two-sided arguments on punishing cheaters, followed by eight short arguments on foreign aid, followed by four long arguments on animal testing; the next subject might rate four long marijuana arguments, then eight short tuition arguments, then four two-sided foreign aid arguments. Their instructions, repeated at the outset of the rating task for each argument type (i.e., three times), were:

In this section, we will ask you to read a set of arguments and tell us how WEAK or STRONG you believe each argument is. Imagine that you have been asked to explain the issue to other students. Considering each argument carefully and objectively, tell us whether it would be a strong or a weak argument to include in your explanation. REMEMBER: whether you agree or disagree with the

⁴ These formed very strong scales with reliability coefficients above .9.

⁵ Attitude position must be folded at the scale midpoint before computing these correlations, since we are testing whether attitude strength is independent of attitude extremity. The sixteen correlations between position and strength for the eight issues at two time points average .28, with a maximum of .41 for tuition raises at time 1.

conclusion of an argument is not the same thing as the degree to which you think the argument is weak or strong.

These instructions were designed to reduce disconfirmation biases in the strength judgment task by pushing for evenhanded treatment. Participants are asked to imagine that they will be talking about the issue with other students, which increases accountability, which has been found to increase accuracy motivation (see Kunda 1990; Taber and Lodge 2006). The instructions also explicitly tell participants (three times in the course of the study) that agreement with an issue is not the same as argument strength. Despite these rather heavy-handed promotions of accuracy, however, we expect that many subjects will not treat the arguments even-handedly.

For half of the arguments rated, the strength-rating screen was immediately followed by a “Stop and Think” screen asking respondents what they were “thinking about when [they] rated the strength of the last argument.” A separate paper form was provided on which to list their thoughts. This procedure measures the content of thoughts in response to arguments and allows us to test for active counterarguing.

After completing the argument strength rating tasks for the three argument types, participants answered a standard set of demographic questions and 17 political knowledge items (e.g., “Name the office held by Antonin Scalia.”). General knowledge or sophistication was measured as the number of correct responses to these 17 questions.

Finally, the same attitude battery was used to collect post-test measures of attitude position and strength on the eight issues.⁶ The most direct measure of attitude change is

⁶ All post-test attitude scales, with the exception of funding of unpopular groups (again), were reliable (all α 's > .7).

the difference between attitude position measured before and after the information processing tasks. This $t_2 - t_1$ difference in attitude position was computed as follows:

Our attitude change measure must increase with movement toward the upper *and* lower bounds of the scale. That is, with a scale that ranges [0, 1] with a true neutral of .5, the same amount of attitude polarization occurs when a t_1 score of .7 is followed by a t_2 score of .9 as when a t_1 score of .3 is followed by a t_2 score of .1. To capture movement toward and away from both extremes, we calculate the change measure as follows.

Change toward the closer scale endpoint is computed as $|t_2 - t_1|$. Change away from the closer endpoint is computed as $-|t_2 - t_1|$. This yields a simple measure of attitude change with the following desirable properties: the $t_2 - t_1$ difference is zero for no attitude change; the difference increases positively for polarization; and it increases negatively for moderation or persuasion. Since the original attitude measures were scaled to [0, 1], this attitude change variable has a range of [-1, .5].

Results

The Attitude Congruence Bias

H1 claims that people will judge counter-attitudinal arguments to be weaker than pro-attitudinal ones. Figure 2 displays average strength ratings (scaled to [0,1]) for the arguments broken down by argument type (short, long, and two-sided), prior position on the issue (supporters or opponents, split at scale neutral point), sophistication of the participant (low or high general political knowledge, by median split), and strength of prior attitude on the issue (weak or strong, by median split). Dark bars represent average strength ratings for pro arguments, light bars con arguments; the first pair of bars in each

pane show the responses of opponents of the issue, and the second pair show responses of supporters (all subgroup Ns exceed 22). The prior attitude bias is indicated wherever we see higher ratings for congruent than incongruent arguments. In other words, we expect supporters to rate pro arguments more highly than they rate con arguments (with the opposite pattern for opponents). Quite clearly, our sample rated the arguments that were congruent with their priors as stronger than incongruent arguments, and this bias appears across all but one subgroup (weak supporters' ratings of long arguments show no bias).

[Figure 2 and Table 2 about here]

To test for attitude congruence bias we regressed argument strength ratings on prior attitudes, as measured by the 4-item scale for pre-test attitude position described above (scaled to $[0,1]$), controlling for sophistication and strength of prior attitudes (see Table 2).⁷ The dependent variable in this regression was computed as the sum of the ratings of pro arguments minus the sum of the ratings of con arguments, scaled to $[0,1]$. Prior attitude significantly predicts bias in argument strength ratings, though marginally so for short and two-sided arguments, and this bias is moderated by strength of priors for all types of arguments. As expected, people are biased by their prior attitudes when they consider policy arguments, especially when these prior attitudes are strong. We did not, however, find any evidence that sophisticates are more prone to this simplest type of bias than are the relatively uninformed.

The Disconfirmation Bias

Within our experimental design, we have two tests of the disconfirmation hypothesis, which posits that participants will actively counter-argue challenges to their

⁷ Recall that attitude position and strength are independent attitude dimensions in our data, as we would expect theoretically and from prior empirical research.

priors, especially when they have strong prior attitudes (motive) and when they are knowledgeable (opportunity). First, we recorded the time it took participants to rate each argument, where longer times indicate deeper processing and indirectly suggest counter-arguing when in response to challenging arguments. Second, we asked participants to list their thoughts immediately following half of the argument rating trials, and we can analyze the content of these thoughts for direct evidence of counter-arguing.

[Figure 3 about here]

Figure 3 displays the average rating times for arguments, broken down by argument type, knowledge, attitude strength, and prior position.⁸ Effortful disconfirmation processes should be indicated by longer rating times for arguments that are incongruent with the respondent's prior position, and this is what we find for knowledgeable participants and those with strong priors. These sophisticated and motivated subjects take 5-8 seconds longer on average to rate inconsistent arguments than to rate consistent arguments, an increase of more than 30%. By contrast, there is no clear pattern for low knowledge participants or those with weak priors, who on average take about the same time to rate consistent and inconsistent arguments. Indeed, rating times for incongruent arguments were significantly longer as compared to congruent arguments among sophisticates across all argument types (short arguments: $t(48)=5.19$, $p<.001$; long arguments: $t(50)=2.87$, $p=.003$; two-sided arguments: $t(49)=3.32$, $p<.001$; all one-tailed tests) and among those with strong prior attitudes for short and two-sided arguments (short: $t(71)=2.97$, $p=.002$; long: $t(77)=0.71$, ns; two-sided: $t(77)=3.13$, $p=.001$). As

⁸ Since the comparison is within subjects (rating time for inconsistent arguments minus time for consistent arguments), we control for individual differences in reading times.

expected, these comparisons were consistently insignificant for unsophisticates and those with weak priors.

Longer rating times indicate deeper processing, and it is not surprising that knowledgeable subjects and those with strong priors are more likely to expend the effort, but what exactly are they doing with this extra time? Analysis of the thoughts listed by participants immediately after rating half of the arguments can answer this question.⁹ Following a coding scheme adapted from Edwards and Smith (1996; Taber and Lodge 2006), we divided all thoughts into three basic categories: *affective statements* (e.g., “I just don’t like this argument”), *new information* (e.g., “this argument doesn’t take into account the aggressiveness of some campus groups”), and *comments* about the source or evidence (e.g., “everyone knows that the *NY Times* is a liberal newspaper”). Each statement was also coded as denigrating or bolstering the argument. We expect that most comments in response to attitude-congruent arguments will be bolstering, while those in response to incongruent arguments will be denigrating, especially for sophisticates and those with strong priors.

[Figure 4 about here]

Figure 4, which shows the average number of categorized thoughts per argument, broken down by argument type, sophistication, and strength of priors, confirms this expectation, at least for short and long arguments. Note first that participants averaged 2 thoughts per short argument, 3.5 thoughts per long argument, and 2.9 thoughts per two-sided argument. But there was great variation: sophisticates and those with strong priors

⁹ Participants listed their thoughts for four short arguments, two long arguments, and two two-sided arguments, balanced across pro and con. They listed their thoughts immediately after rating each argument.

tended to have more thoughts, especially denigrating thoughts in response to incongruent arguments. A 2 (high and low sophistication) by 2 (weak and strong priors) by 2 (congruent or incongruent argument) by 2 (denigrate or bolster) by 3 (short, long, or two-sided argument) ANOVA on the number of thoughts yielded powerful main effects for sophistication, $F(1,47)=9.66$, $p=.006$, argument congruence, $F(1,47)=23.03$, $p<.001$, response type, $F(1,47)=17.21$, $p=.001$, and argument type, $F(2,47)=7.45$, $p=.004$. Strength of priors did not produce a significant main effect. Follow-up contrasts, which we do not detail to conserve space, confirm the patterns in Figure 4: sophisticates had significantly more thoughts than unsophisticates; incongruent arguments spurred more thoughts than did congruent ones; there were more denigrating than bolstering thoughts; and long arguments caused more thoughts than did two-sided arguments, which caused more thoughts than short arguments.

More important for the disconfirmation hypothesis, the interaction of argument congruence and response type was highly significant, $F(1,47)=140.99$, $p<.001$, such that incongruent arguments caused significantly more denigrating thoughts while congruent arguments caused bolstering thoughts (planned contrast: $F(1,18)=25.17$, $p<.001$). Next, we found three-way interactions of argument congruence and response type with sophistication $F(3,47)=17.01$, $p<.001$, and strength of prior, $F(3,47)=5.45$, $p=.008$. Sophisticates more than unsophisticates ($F(1,18)=23.91$, $p<.001$) and those with strong priors more than weak ($F(1,18)=12.70$, $p=.002$) denigrated challenging arguments and bolstered supportive ones. Interestingly, though it appears in Figure 4 that the greatest bias occurred for long arguments, this interaction was not significant. Though there were

more thoughts in general for one-sided arguments, there was not more *bias* in the content of thoughts. There was, in fact, considerable bias for all argument types.

Clearly respondents in our experiment sought to disconfirm challenging arguments. They spent significantly more time pondering these arguments, and they generated more counterarguments when presented with evidence that challenged their priors. Most importantly, those with the strong priors and sophistication that might motivate and enable disconfirmation processes were most susceptible to these biases. Despite our best efforts to motivated accurate and even-handed treatment of the evidence, our participants displayed powerful tendencies to accept congenial arguments with little scrutiny while quite actively rejecting the arguments that challenged their prior positions on the issues.

Polarization

Biased processing of political arguments should polarize attitudes. That is, attitudes on these issues should become more extreme as a consequence of the attitude congruence and disconfirmation biases, even when the overall set of arguments is balanced between pro and con. Polarization, moreover, should be stronger for sophisticates and those with strong prior attitudes as a consequence of their more biased information processing. Psychologists have found it difficult to capture this theoretical expectation in the lab, in part because of weak measurement of pre and post attitudes and in part because their stimulus materials are usually not very affectively engaging. We have addressed both of these weaknesses in the current study (see also Taber and Lodge, 2006).

Averaging across all participants, attitudes in our study changed very little from t_1 to t_2 . Indeed, as captured by the $t_2 - t_1$ attitude change variable described earlier, there was essentially no change for short and long arguments and only slight polarization for two-sided arguments. But this grand average conceals substantial and theoretically meaningful variation. Table 3 contrasts average attitude change for high versus low sophisticates, participants with weak versus strong attitudes, and those who showed high versus low processing bias. Recall that attitude change greater than zero indicates polarization, while change below zero indicates moderation (and shades into persuasion the more negative it becomes).

[Tables 3 and 4 about here]

We used two measures of processing bias as predictors of polarization: (1) an attitude congruence bias measure, computed as the average ratings for attitude-congruent arguments minus the average ratings for attitude-incongruent arguments; and (2) a disconfirmation bias measure, computed as the number of attitude-consistent thoughts (bolstering thoughts for pro arguments and denigrating thoughts for con arguments) generated in the thoughts listing task divided by the total number of thoughts. Median splits on these measures distinguish high and low bias groups in Table 3.

A regression of sophistication, attitude strength, and two measures of processing bias – congruence bias in argument ratings and disconfirmation bias – on t_2-t_1 attitude change is reported in Table 4. Across all three argument types, bias in processing significantly predicts attitude movement in the direction of polarization. Neither attitude strength nor sophistication consistently drive polarization in these regressions, however, which is puzzling given the group differences shown in Table 3 for attitude strength. But

since the bias measures are themselves strongly predicted by strength of prior attitude, and since the bias variables are causally more proximate to attitude change, we should expect bias to mask the influence of strength of prior in these regressions. Attitude strength, but not sophistication, emerges as a significant predictor of attitude change for long and short arguments when the bias variables are removed from the regression equations. This is consistent with our theoretical expectation about the role of attitude strength in promoting bias and thereby promoting polarization.

In short, those who process arguments in a biased fashion and those with strong prior attitudes are more likely to polarize than their contrast groups. But this does not establish that anyone actually polarized. To do so, we need to show that the attitude changes for these groups are significantly greater than zero, not just that variation in attitude change is predictable in the expected direction. We ran *t*-tests for the difference from zero of the positive group means in Table 3. Contrary to expectation, but predictable from the regressions just reported, attitude change for sophisticates was never significantly greater than zero. In these data, sophistication does not predict polarization. Participants with strong priors, however, did significantly polarize after evaluating the arguments: for long arguments, $t(81)=4.55, p<.001$; short arguments: $t(81)=4.72, p<.001$; two-sided arguments: $t(81)=2.07, p=.02$, all one-tailed tests. Even more impressive, those who showed the most bias in the argument rating task also consistently polarized: for long arguments, $t(81)=5.31, p<.001$; short arguments: $t(81)=1.95, p=.03$; two-sided arguments: $t(81)=5.73, p<.001$, all one-tailed tests.

Argument Type

In this study, we sought to test two expectations about how type of argument might influence processing bias, viewing them as potential boundary conditions for motivated reasoning: first, we predicted that longer, more complex one-sided arguments would generate more motivated bias than shorter, simpler arguments. Indeed our own previous studies, which used complex and evocative arguments like the long arguments in the current study, have generally uncovered stronger biases than most experiments by social psychologists, which used simpler declarative statements like our short arguments. Second, we predicted that two-sided arguments, which embed a weak counterargument, would generate less motivated bias than one-sided arguments of equal complexity. This expectation was based on suggestive evidence from social psychology that such arguments, which are found frequently in real world discourse, may have greater credibility than one-sided arguments.

The analyses we have already reported, however, do not support either hypothesis. We found solid evidence of attitude congruence bias across all three argument types; indeed, for those with strong prior attitudes, attitude congruence bias was larger for short than for long arguments and of comparable size for two-sided arguments (see the Prior x Strength interactions in Table 2). We also found disconfirmation biases for all three argument types. The ANOVA results on the thoughts our participants listed shows more thoughts for long arguments than short or two-sided arguments, but *not* more biased thoughts. That is, the interaction of argument type (long, short, two-sided) with response type (denigrating or bolstering arguments) and the three-way interaction of argument type, response type, and argument congruence were insignificant. This shows

that the degree of bias in thoughts listing, which is indicated when attitudinally incongruent arguments spark denigration and congruent arguments cause bolstering, does not differ significantly across the three argument types. Finally, we found polarization across all argument types, with nary a hint of the expected greater bias for long arguments.

To confirm the failure of our argument type hypotheses, we computed *t*-scores for the difference in bias between long arguments on the one hand and short and two-sided arguments on the other. Using the two measures of processing bias described in the last section – attitude congruence bias and disconfirmation bias – we found slightly more bias for short than for long arguments, which is the opposite of our expectation, and no significant difference between long and two-sided arguments. In short, this study found no evidence that motivated biases are contingent on argument length, complexity, or credibility.

Discussion

The literatures on political attitudes and public opinion have long puzzled over why some attitudes are more persistent than others. Clearly the question has important ramifications for the functioning of democracy and for the individual rationality of citizens. One theoretical explanation, the one explored in this paper, argues that some political attitudes are *actively defended* in the course of information processing. That is, citizens process counter-attitudinal arguments or evidence differently than they process attitude-consistent arguments or evidence, holding the arguments they don't like to a higher standard.

This paper reports strong evidence that people cannot escape the pull of their prior beliefs when considering policy arguments. Arguments that are congruent with priors are perceived to be stronger than incongruent ones. A student who believes that cheaters should be punished will reliably find the arguments *for* stronger university punishment to be more compelling than those against. Moreover, this student will actively denigrate arguments from the other side, forming counter-arguments that allow her attitudes to persist and even strengthen in the face of counter-evidence. In terms of our research hypotheses, we found strong and consistent evidence of an *attitude congruence bias*, a *disconfirmation bias*, and a resultant *attitude polarization*. Our design allows us to show that this polarization is a direct consequence of attitude congruence and disconfirmation biases.

We also expected that these biases would be moderated by sophistication and strength of prior attitudes, such that higher levels of each should lead to greater bias and polarization. While we found consistent evidence that attitude strength moderates all forms of bias and polarization, sophistication was an important moderator of disconfirmation biases, but not attitude congruence bias and not polarization. Though we did not predict this pattern of results for sophistication, it does in fact make some *post hoc* sense: it takes very little sophistication to feel that attitude-congruent arguments are stronger than those that challenge one's priors; the active generation of counter-arguments on the other hand should require more sophistication.

Finally, in an effort to explore potential boundary conditions for motivated reasoning, we speculated that argument type might moderate processing bias. Two possibilities were considered: (1) longer, more complex arguments might generate greater

bias than short, direct statements and (2) one-sided arguments might spark more biased reactions than two-sided arguments, which may be perceived as more credible and therefore be more persuasive. Neither of these possibilities was at all supported by our experiment, and we must conclude that motivated reasoning in these experiments was robust to these variations in argument style.

Despite our best efforts to motivate accurate and even-handed treatment of the evidence, our participants displayed powerful tendencies to accept congenial arguments with little scrutiny while actively rejecting the arguments that challenged their prior positions on the eight issues in our study. Most troubling from a normative perspective, our work here and elsewhere (Taber and Lodge 2006) shows that people may become more extreme in their beliefs as they read balanced sets of pro and con arguments. The extent to which they polarize is contingent upon how biased their information processing is, which depends on how strongly they feel about the given issue. Most important, this experiment shows that attitude polarization is attributable to the degree of bias in information processing; it is in fact a direct consequence of motivated reasoning.

But all reasoning is motivated. The same feelings that produce biased information processing motivate citizens to participate in politics. Moreover, the biases we document are only problematic to the extent that the prior attitudes involved were not generated through a normatively acceptable process. This does not seem to us to resolve the normative problem, however, since so many politically consequential attitudes are born of whim and fancy and justified on the flimsiest of grounds (attitudes on gay rights, abortion, or almost any form of prejudice, e.g.). Worse, to the extent that these biases are present from the first moment one evaluates a political object and operate *automatically*,

they would seem to militate against any normatively acceptable process for creating and updating attitudes. Built into our cognitive and emotional processing systems are mechanisms that motivate and enable political action. The irony of motivated reasoning is that these same mechanisms undermine our ability to rationally respond to new and challenging information.

References

- Abelson, R. P., & Prentice, D.A. (1989). Beliefs as possessions: A functional perspective. In A.R. Pratkanis, S.J. Breckler, & A.G. Greenwald (Eds.), *Attitude structure and function* (pp. 361-381). Hillsdale, NJ: Erlbaum.
- Chong, D., & Druckman, J. N. (2007). Framing theory. *Annual Review of Political Science*, 10, 103-26.
- Ditto, P. H., & Lopez, D. F. (1992). Motivated skepticism: Use of differential decision criteria for preferred and nonpreferred conclusions. *Journal of Personality and Social Psychology*, 63(4), 568-584.
- Ditto, P. H., Scepansky, J. A., Munro, G. D., Apanovitch, A. M., & Lockhart, L. K. (1998). Motivated sensitivity to preference-inconsistent information. *Journal of Personality and Social Psychology*, 75, 53-69.
- Edwards, K., & Smith, E. E. (1996). A disconfirmation bias in the evaluation of arguments. *Journal of Personality and Social Psychology*, 71, 5-24.
- Evans, J. St. B. T., & Over, D. E. (1996). *Rationality and reasoning*. Hove: Psychology Press.
- Fischhoff, B., & Beyth-Marom, R. (1983). Hypothesis evaluation from a Bayesian perspective. *Psychological Review*, 90, 239-260.
- Gerber, A., & Green, D. P. (1998). Rational learning and partisan attitudes. *American Journal of Political Science*, 42(3), 794-818.
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108(3), 480-498.

- Lebo, M., & Cassino, D. (2007). The aggregated consequences of motivated reasoning. *Political Psychology*, 28, 6.
- Lodge, M., & Taber, C. S. (2000). Three steps toward a theory of motivated political reasoning. In A. Lupia, M. McCubbins, & S. Popkin (Eds.), *Elements of reason: Cognition, choice, and the bounds of rationality* (pp. 183-213). New York: Cambridge University Press.
- Lodge, M., & Taber, C. S. (2005). Implicit affect for political candidates, parties, and issues: An experimental test of the hot cognition hypothesis. *Political Psychology*, 26(3), 455-82.
- Lord, C., Ross, M., & Lepper, M. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, 27, 2098-2109.
- Matheson, K., & Dursun, S. (2001). Social identity precursors to the hostile media phenomenon: Partisan perceptions of coverage of the Bosnian conflict. *Group Processes and Intergroup Relations*, 4, 117-126.
- Meffert, M., Chung, S., Joiner, A. J., Waks, L., & Garst, J. (2006). The effects of negativity and motivated information processing during a political campaign. *Journal of Communication*, 56 (1), 27-51.
- Petty, R. E., & Krosnick, J. (Eds.) (1995). *Attitude strength: Antecedents and consequences*. Hillsdale, NJ: Lawrence Erlbaum.
- Petty, R. E., & Wegener, D. T. (1998). Attitude change: Multiple roles for persuasion variables. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (pp. 323-390). Boston: McGraw-Hill.

- Redlawsk, D. (2002). Hot cognition or cool consideration: Testing the effects of motivated reasoning on political decision making. *Journal of Politics*, 64, 1021-1044.
- Rudolph, T. J. (2006). Triangulating political responsibility: The motivated formation of responsibility judgments. *Political Psychology* 27(1), 99-122.
- Taber, C. S. (2003). Information processing and public opinion. In D. O. Sears, L. Huddy, & R. Jervis (Eds.), *The handbook of political psychology* (pp. 433-476). New York: Oxford University Press.
- Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, 50(3), 755-69.
- Taber, C. S., Lodge, M., & Glather, J. (2001). The motivated construction of political judgments. In J. Kuklinski (Ed.), *Citizens and politics: Perspectives from political psychology* (pp. 198-226). New York: Cambridge University Press.

Table 1: Sample Arguments

Short Argument	Long Argument	Two-sided Argument
Smoking marijuana is a stepping stone to harder drugs, including cocaine and heroin.	Smoking marijuana is a stepping stone to harder drugs, including cocaine and heroine. Most serious addicts began by smoking pot. Experience shows that serious drug addiction usually starts with marijuana.	Smoking marijuana is a stepping stone to harder drugs, including cocaine and heroin. Some argue that legalizing marijuana would at least help move drug abusers from hard drugs to marijuana. But experience shows that serious drug addiction usually starts with marijuana.
Freedom of speech is a cherished democratic value that is meant to protect everyone, not just those we like or agree with.	Freedom of speech is a cherished democratic value that is meant to protect everyone, not just those we like or agree with. Tolerance of others' views is required in a democracy. Allowing individual students to silence those they dislike by refusing to pay the mandatory student activity fee goes against the very values this country stands for.	Freedom of speech is a cherished democratic value that is meant to protect everyone, not just those we like or agree with. Some might argue that the refusal to finance the freedom of speech of others is itself an expression of free speech. Nevertheless, allowing individual students to silence those they dislike by refusing to pay the mandatory student activity fee goes against the very values this country stands for.

Table 2: Attitude Congruence Bias in Argument Ratings

	Long Arguments	Short Arguments	Two-Sided Arguments
Prior Attitude	.096 ^a (.034)	.092 ^c (.051)	.036 ^c (.022)
Sophistication	.024 (.044)	.037 (.099)	.052 (.059)
Strength of Prior	-.110 ^b (.054)	-.207 ^b (.081)	-.095 ^b (.044)
Prior x Sophistication	.034 (.073)	-.058 (.161)	.011 (.095)
Prior x Strength	.170 ^c (.096)	.418 ^a (.135)	.215 ^a (.069)
Constant	.445 (.026)	.465 (.039)	.440 (.024)
R^2	.314	.330	.313
Mean Squared Error	.072	.117	.075
N	164	164	164

Note: Cells report unstandardized regression coefficients and standard errors for the influence of prior attitude (supporters above .5) and two moderators on strength ratings (average ratings of pro arguments minus average ratings of con arguments). All variables are coded [0, 1]. Columns report separate OLS regressions for the three argument types. Level of significance is indicated as follows: ^a <0.01, ^b <0.05, ^c <0.10.

Table 3: Attitude Polarization by Sophistication, Strength of Prior, and Bias

	Long Arguments		Short Arguments		Two-Sided Arguments	
	Mean	SD	Mean	SD	Mean	SD
All Subjects (N=164)	-.005	.275	-.001	.226	.030 ^b	.191
Sophistication						
Low (N=85)	.013	.227	-.021	.254	.042 ^b	.160
High (N=79)	-.025	.320	.021	.189	.018	.220
Strength of Prior						
Weak (N=82)	-.111	.300	-.083	.254	.013	.212
Strong (N=82)	.101 ^a	.200	.081 ^a	.155	.048 ^b	.166
Congruence Bias						
Low (N=82)	-.133	.275	-.051	.216	-.048	.177
High (N=82)	.123 ^a	.210	.049 ^b	.226	.109 ^a	.172
Disconfirmation Bias						
Low (N=77)	-.055	.287	-.065	.251	.005	.215
High (N=87)	.039 ^c	.259	.048 ^a	.191	.062 ^a	.165

Note: Cells report means and standard deviations for the t_2-t_1 attitude change variable. All positive means significantly greater than zero in a one-tailed t-test are indicated as follows: ^a <0.01, ^b <0.05, ^c <0.10.

Table 4: Attitude Change as a Function of Sophistication, Attitude Strength, and Motivated Bias

	Long Arguments	Short Arguments	Two-Sided Arguments
Sophistication	-.013 (.091)	.132 (.098)	-.017 (.064)
Strength of Prior	.153 ^c (.091)	.074 (.101)	-.063 (.065)
Congruence Bias	1.265 ^a (.282)	.383 ^b (.193)	.876 ^a (.165)
Disconfirmation Bias	.281 ^c (.148)	.189 ^c (.106)	.146 ^b (.063)
Constant	-1.007 (.206)	-.564 (.157)	-.407 (.116)
R^2	.198	.123	.202
Mean Squared Error	.258	.214	.155
N	149	117	143

Note: Cells report unstandardized regression coefficients and standard errors for the influence of predictors on t_2 - t_1 attitude change. Columns report separate OLS regressions for the three argument types. Level of significance is indicated as follows: ^a <0.01, ^b <0.05, ^c <0.10.

Figure 1: The Experimental Design

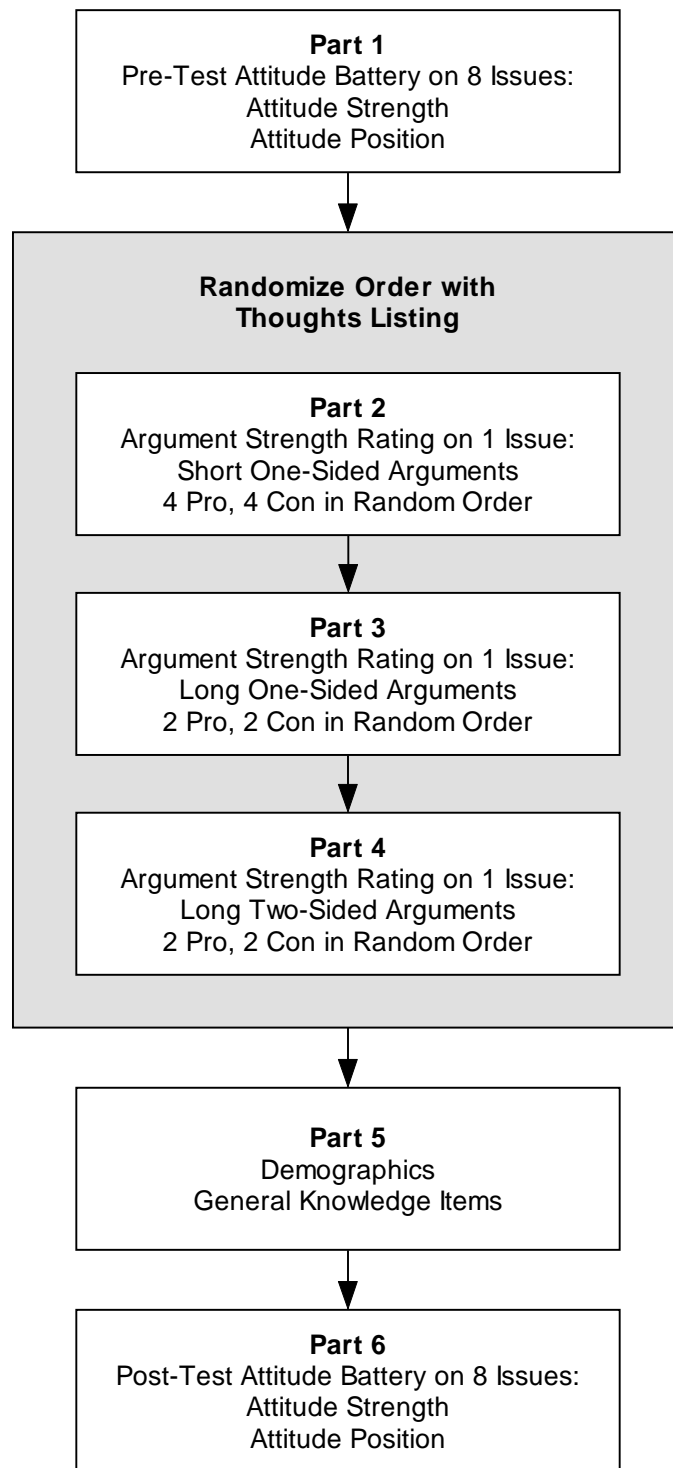


Figure 2: Argument Strength Ratings by Prior Position

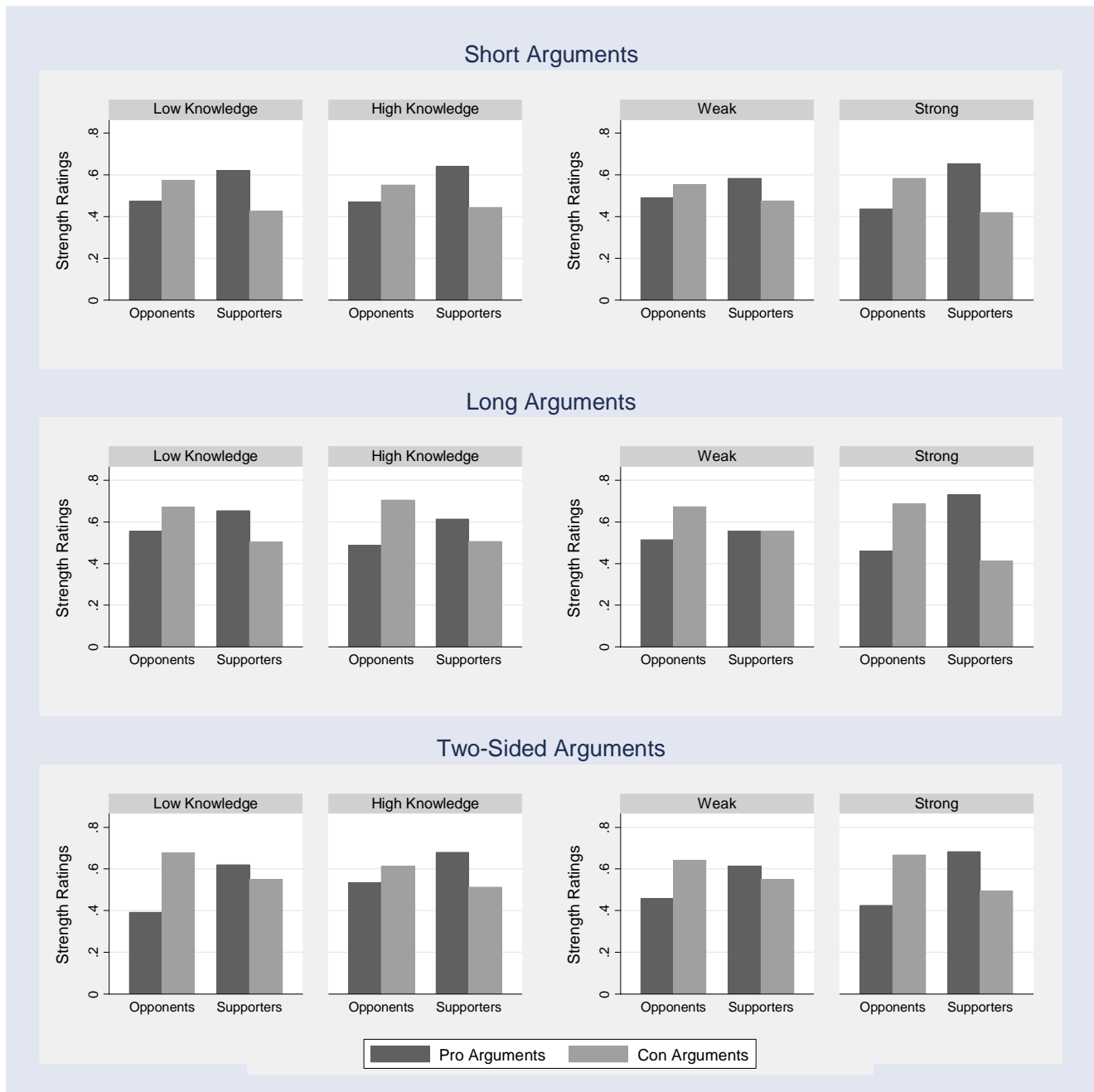


Figure 3: Argument Rating Times by Congruence, in Seconds

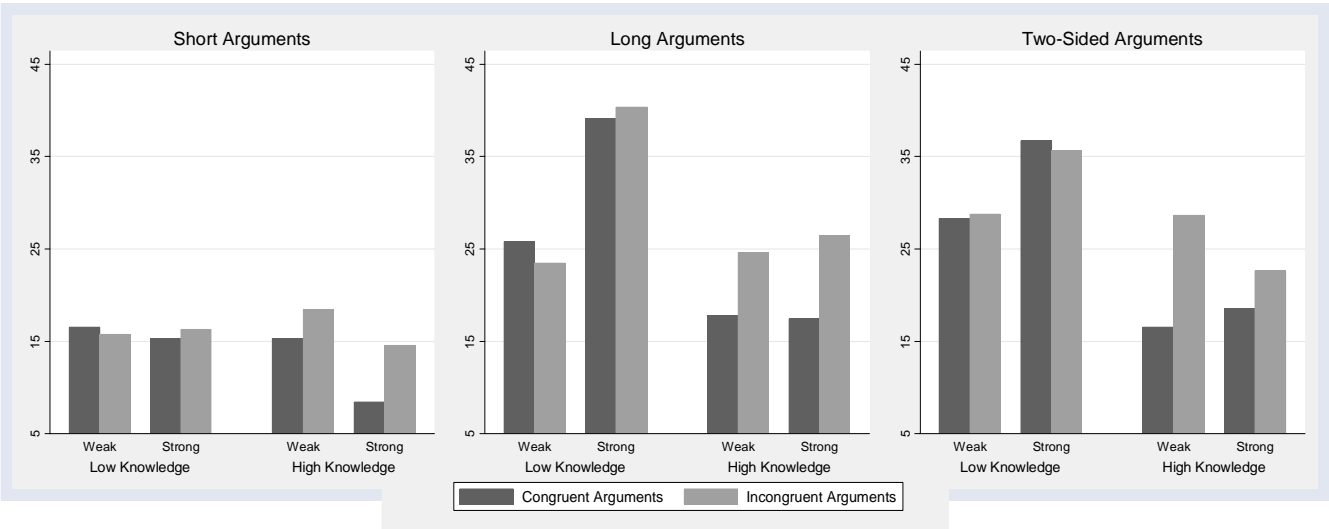


Figure 4: Mean Number of Thoughts by Type and Congruence

