Will Google destroy western democracy? Bias in policy problem solving

Matthew W. Easterday, Vincent Aleven, Richard Scheines, Sharon M. Carver
Human-Computer Interaction Institute, Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh PA, 15213-3891

Abstract: Democracy requires students to choose policy positions based on evidence, yet confirmation bias prevents them from doing so. As a preliminary step in building a policy reasoning tutor, this study identifies where bias occurs during the search and analysis of evidence in a policy reasoning task. 60 university students played an on-line game in which they chose which of four policies would increase school performance. The between-subjects design compared a free search group who searched for evidence in google-like environment, to a sequential presentation group who read all available evidence, and manipulated whether the evidence confirmed or disconfirmed students' prior beliefs. The study measured the impact on students' evidence-based recommendations, their change in beliefs, and their recall of the evidence. Results showed that students did not cherry-pick evidence nor discount disconfirming evidence. However, students' extreme confidence in their initial beliefs usually prevented them from changing position, and they mistakenly recalled the evidence as confirming their beliefs. The results suggest that a policy tutor should focus on evidence synthesis and making recommendations based on *explicit* evidence.

Keywords: policy, causal reasoning, motivated skepticism, evidence evaluation

It is axiomatic that democracy depends on an active, engaged citizenry. Having shown that citizens do not reason rationally on the basis of evidence about policy issues such as global warming or financial regulation, some psychologists have argued that leaders should appeal more to emotion or better frame the issues [1,2]. While more persuasion might win elections, it will do little to promote an active, engaged citizenry that demands better policy. The alternative is civic education [3]. As a preliminary step toward a cognitive tutor for civics, this study examines where bias occurs during a *policy reasoning* task [4-6].

A policy reasoning task, like deciding whether decreasing classroom size will increase school performance, requires reasoners to decide whether a policy will lead to the desired outcome on the basis of evidence. Unfortunately, psychologists have shown a consistent pattern of confirmation bias, where the reasoner's prior beliefs overwhelm impartial evaluation of evidence [7-11]. On emotional topics such as the death penalty and gun control, psychologists have even shown that partisans from opposite sides can each become more convinced of their original position after seeing the same set of evidence [12]. Bias arises both during search [13,14] and during analysis of information [15-18, 8].

Previous work by the authors [19] argued that causal diagrams might allow us to create policy reasoning tutors, but like many studies on policy reasoning, used a task in which students evaluate arguments [4,13] as opposed to empirical evidence [12], focusing on analysis [4] rather than search [13,14]. The purpose of this study was to test where bias and errors occur during a policy reasoning task that requires both search and analysis in order to provide a more authentic problem domain for a policy tutor

[20]. In this task, students had to recommend which of four policies would increase school performance by searching and evaluating empirical evidence.

The study examined the effect of students' ***search*** for evidence and the congruence of evidence with students' ***prior beliefs*** on their ability to make ***evidence-based recommendations***. For example, if a student believes that decreasing classroom size doesn't increase school performance, then she might not recommend decreasing classroom size even if she is given evidence to the contrary. The study manipulated prior belief by providing evidence mostly congruent with the student's belief about one policy (e.g., decreasing class size), and another set of evidence mostly incongruent with the student's belief about a second policy (e.g., requiring teachers to have masters degrees). To examine *search,* the study compared a free search group who searched for evidence in a simulated google environment (the *google* group), to a sequential presentation group who read every piece of information in a fixed order, specifying their beliefs after each piece of evidence (the *1x1* group). The study measured the ***evidence*** read by each student, students' ***confidence shifts*** in their beliefs after reading the evidence, and students' ***recall of evidence.*** Based on the consistent pattern of confirmation bias in previous research, we predicted bias at all stages of processing:

H1: Biased search Google students will search for reports in a biased manner.
H2: Biased evaluation Students will shift their beliefs more to confirming than disconfirming evidence.
H3: Biased synthesis Students will overestimate the amount of congruent evidence read.
H4: Biased recommendation Students will make recommendations more consistent with their beliefs than with the evidence.

Method

Population and setting

60 University students were recruited through an on-line subject database to participate in the study. Participants had a median age of 21 years, 95% were bilingual or native English speakers, and 92% used the internet at least once a day. Participants completed the study over the internet.

Procedure

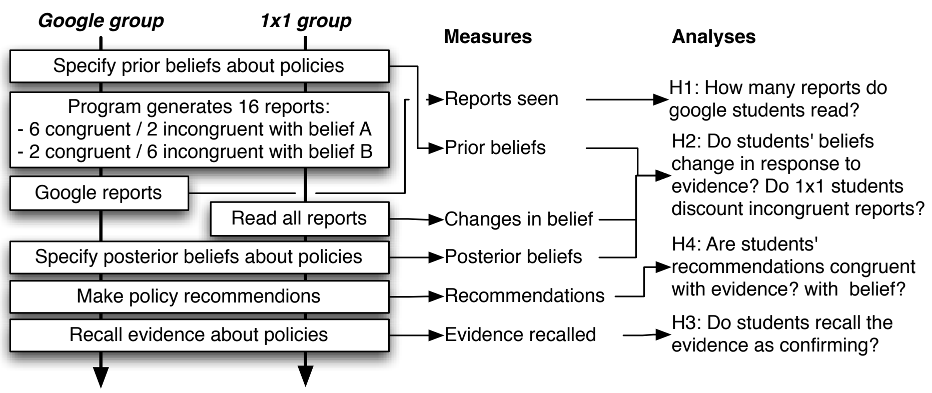
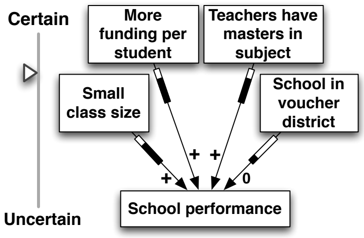


Figure 1. Experimental procedure, measures and analyses.

Students played a computer game in which they assumed the role of policy analysts. Their goal was to determine whether four different policies: reducing class size, increasing teacher qualifications, increasing funding, or providing vouchers, would increase school performance. At the beginning of the game, students specified their prior beliefs about each policy: whether the policy would have a positive, negative or negligible effect on school performance, and their certainty in their belief (Figure 2a).



Example report:
Dr. Jones, a professor of educational policy at Harvard University, studied high schools from 12 different states on many dimensions including class size. The schools were evaluated on how well their students performed on the NAEP test of mathematics. Mr. Jones found that schools with smaller classes performed no differently than schools with larger classes on the NAEP. When asked about the implications of this research, Dr. Jones implied that more work like this needs to be done in order to fix America's schools.

Figure 2a. Belief graph (left) A student specified his belief about each policy by toggling a +/0/- button on a causal arrow and moving a slider to indicate his certainty in that belief. **2b. Example report (right)** summarizes an observational study that shows no effect of smaller class size on school performance, which would be incongruent with a prior belief that smaller class size increases performance.

In both groups, the game identified the two policies about which the student had the strongest beliefs and generated two sets of evidence in the form of one-paragraph descriptions mimicking newspaper reports (Figure 2b). One set of 8 reports was *mostly* congruent (6 congruent, 2 incongruent) with one of the student's beliefs (e.g., about class size), and a second set of 8 reports was *mostly* incongruent with a second policy belief (e.g., about teacher qualifications). Half the reports summarized observational studies, and half case studies. The game then randomly assigned students to either the google group which searched freely for reports (Figure 3), or *1x1* group which read every report and specified the change in their beliefs after each report.

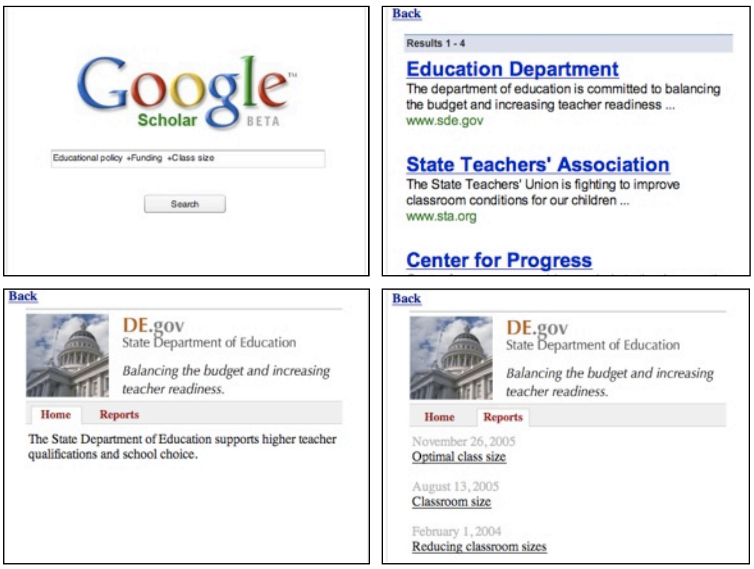


Figure 3. Google group screen shots. Google students searched simulated web pages generated by the game to find the reports. Google students saw a home page stating the organization's biases (bottom left), before reaching the list of reports (bottom right) leading to an individual report like in Figure 2b. The 1x1 students saw only the reports, but had to read each report.

After reading the reports, all students specified their final beliefs about whether each policy would affect school performance, made their recommendations about which policies should be implemented (yes/no), and specified how much evidence they thought they read about each policy (number of reports seen about each policy and the proportion of reports indicating the policy would work).

Results

H1: Biased Search. To test the first prediction that google students would search in a biased manner, the first analysis compared the reports read by each group and the proportion of the reports read that were congruent with the students' prior beliefs.

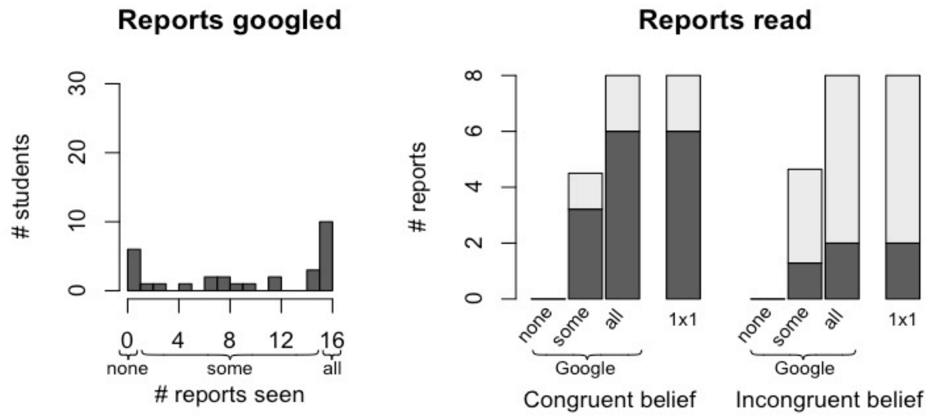


Figure 4a. Reports googled (left) shows that 6 google students read *none* of the reports, 14 read *some* (1-15) reports, and 10 read *all* reports. **4b. Reports read** *(right)* shows the number of congruent (dark gray) and incongruent (light gray) reports read by the 1x1 students and the google students who read all, some, or no reports. For the belief mostly congruent by evidence, the 30 students in the 1x1 group read the same proportion of confirming evidence (M = 0.75, SD = 0) as the 24 google students who read some or all reports (M = 0.71, SD = 0.22), $t(23) = 0.93$, $p > .36$. For the belief mostly incongruent with evidence, the 30 students in the 1x1 group read the same proportion of confirming evidence (M = 0.25, SD = 0) as the 23 google students who read some or all reports (M = 0.26, SD = 0.26), $t(22) = 0.50$, $p > .62$.

About two thirds of the google students read less than all reports (Figure 4a), however these students still read approximately the same proportion of congruent / incongruent evidence as students who viewed all reports (Figure 4b). In other words, the google students did not always search everything, but they didn't search in a biased manner. This finding disconfirms the first prediction that google students will selectively search for evidence that confirms their beliefs.

H2: Biased evaluation. To test the second prediction that students shift confidence in their beliefs more to congruent than to incongruent reports, the game logged students' beliefs about each of the four policies before and after reading the evidence. Analysis of the total shift in belief by each group (Figure 5a) did not show the sort of belief polarization found in other policy reasoning studies. In fact, even for the policy belief for which mostly congruent evidence was provided, students decreased their confidence in their original belief, (recall that, for the belief confirmed by evidence, 2 of the 8 reports conflicted with the student's belief).

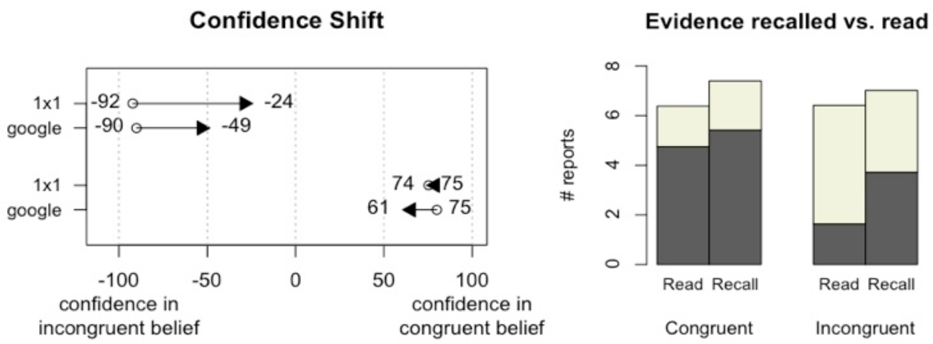


Figure 5a. Confidence shifts (left) shows that the 30 1x1 students decreased their confidence in the incongruent belief by 67% (*SD* = 71%) and in the congruent belief by 1% (*SD* = 40%). The 30 google students decreased their confidence in the incongruent belief by 41% (*SD* = 55%) and in the congruent belief by 19% (*SD* = 59%). A linear mixed model regressed students' shifts in confidence on whether the student was in the google or 1x1 group and whether or not the student's initial belief was congruent with the evidence, with student as a random effect. Students decreased their confidence (b =) 65% more when the evidence was incongruent with their belief, $t(59) = 6.18$, $p < .0001$. Google students did not shift their confidence in the direction of the evidence as much as 1x1 students, (b = 22%), $t(58) = -2.13$, $p < .04$. **5b. Evidence recalled vs. read** (right) students who read at least one piece of evidence about both the congruent and incongruent policy (n = 53) recalled the evidence about the congruent policy accurately (M = -2%, SD = 28%), but recalled the evidence about the incongruent policy as far more confirming than what they actually read (M = 26%, SD = 26%). A t-test showed that the difference (*M = 28%, SD = 35%)* was significant, $t(52) = 5.9$, $p < .0000003$.

Lacking a normative psychological theory of belief updating, we can (imperfectly) define confidence shift as *positive* when it is toward the belief supported by evidence, and *negative* when it is away from the belief supported by evidence. Given this definition, Figure 5a shows that students' confidence shifts more in response to incongruent evidence than to congruent evidence, and that 1x1 students shift confidence more normatively (in the direction of the majority of evidence) than google students (note however that an alternate model of *absolute* confidence shift shows no significant difference between google and 1x1 students).

In the 1x1 condition, the game also solicited students' beliefs after reading each report. Analysis of the 1x1 students' shifts in confidence after each report shows that on average, students shifted in the correct direction by 12% after reading each report (SD = 38%), shifting in the wrong direction only 7% of the time. Students also responded more to incongruent reports (M = 17%, SD = 44%) than to congruent reports (M = 8%, SD = 33%), which a linear regression analysis indicated was significant difference (b = 8.6%, p < .15) accounting for 1% of the variance.

Taken together these results conflict with the hypothesis that students will discount disconfirming evidence. However, students did not behave as normatively as one might desire: their initially high confidence in their prior beliefs meant that their shift in confidence did not always result in a qualitative change in belief (e.g., from believing that smaller classes increase performance, to a belief that class size doesn't matter), and the 1x1 students' data show that they did not shift their confidence more in response to observational studies than to case studies.

H3: Biased synthesis. To test the third hypothesis that students recall the evidence as more congruent with their beliefs than what they actually read, we asked students to specify how many reports they read about each policy, and the percentage of reports that indicated that the policy would work. The results in Figure 5b show that students recall the evidence as far more congruent with their belief than what they actually read, supporting the hypothesis.

H4: Biased recommendations. To test the fourth hypothesis that students make recommendations more congruent with their prior beliefs than with evidence available (or read), we measured students' prior and posterior beliefs and whether they recommended the given policy. The results in Figure 6a show that students' recommendations are more consistent with their posterior beliefs than they are with the evidence, supporting the hypothesis.

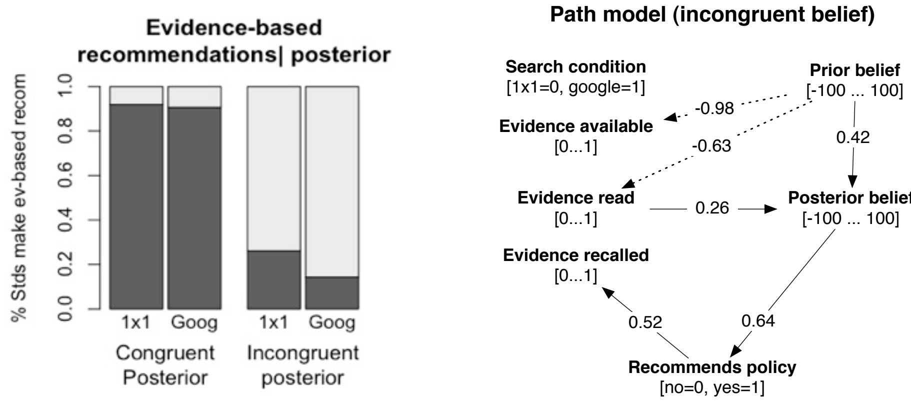


Figure 6a. Evidence based recommendations by posterior belief (left). Students make recommendations congruent with the evidence 91% of the time when their posterior belief is congruent with the evidence (n = 69), but only 20% of the time when their posterior belief is incongruent with the evidence (n = 51), $t(85)$ = 10.9, $p < 2.2\text{e-}16$. **6b. Path model** (right) for the policy belief incongruent with the evidence.

Path Analysis. To understand the relation between beliefs, evidence and recommendations, we analyzed the relations between: (a) the *search condition*, 1x1 or google, (b) the student's *prior belief* about whether the policy had a causal effect, e.g., that they were 50% certain that the policy had a positive effect on school performance before reading the evidence, (c) the percentage of *evidence available* (reports) indicating the policy had a causal effect, (d) the percentage of *evidence read* by the student indicating the policy had a causal effect, (e) the *causal evidence recalled*, (f) the student's *posterior belief*, and (g) whether the student *recommended* the policy.

Table 1. Path model correlations for policy beliefs incongruent with the evidence (n = 60).

	Search	*Prior*	*Evidence*	*Ev. read*	*Ev. recall*	*Post*	*Recommend*	Mean	SD
Search	1.000							0.50	0.50
Prior	-0.091	1.000						81.83	41.54
Evidence	0.076	-0.976***	1.000						
Ev. read	0.250*	-0.628***	0.616***	1.000				0.30	0.14
Ev. recall	0.006	0.165	-0.163	0.145	1.000			0.53	0.28
Post	0.132	0.418***	-0.411***	-0.101	0.384**	1.000		26.83	69.46
Recom.	0.141	0.352**	-0.324**	-0.007	0.523***	0.645***	1.000	0.67	0.48

*$p < .05$ **$p < .01$. ***$p < .001$.

We used the Tetrad program [21, 22] to search among the 221 possible path analytic models consistent with the correlations between the variables (Table 1) and the knowledge that search condition and prior beliefs were set before evidence became available, which preceded the reading of the evidence, which preceded measurement of the evidence recalled, posterior beliefs, and recommendations. The path model (Figure 6b) suggests that students' prior beliefs influence their posterior beliefs, which influence their recommendations. While evidence also influences students' posterior beliefs, it has a weaker effect than prior belief, and students' recall of the evidence merely rationalizes (rather than causes) their final recommendations. A chi-squared test of the deviance of the path model from the observed values (where larger p-values indicate better fit) showed that the predictions of the model did not differ significantly from the observed values, χ^2 (15, n = 60) = 21.78, p > .11.

Discussion

At first glance, the results seem contradictory and only partially consistent with previous literature: no bias was found in students' search or response to evidence, yet the evidence seemed to have little impact on students' final recommendations. In fact the results are consistent with the previous work on bias and, upon closer analysis, show that students are not as biased as the previous work implies. Students essentially respond rationally to any given piece of evidence, increasing their confidence in response to congruent information and decreasing their confidence in response to incongruent information in a relatively symmetric way. However, because the strength of their prior beliefs is so high, each individual piece of evidence affects the prior belief only slightly, like drips of water on stone. Furthermore, if students' beliefs represent only a single overall *impression* [23-24, 14] and do not catalog the evidence read, then when asked to recall information, students can only recreate an answer based on their overall confidence. In this task, students do not so much process information in a biased manner as begin from an extreme position–and, with no clear picture of how the mass of evidence compares to belief, students do not recognize the inconsistency of their beliefs with the evidence. Of course, because they cannot articulate the evidence supporting their initial beliefs, we cannot say whether these initial beliefs are warranted.

These findings have several implications for developing a policy tutor:

1. A policy tutor should focus first on evidence synthesis (as opposed to search or comprehension). While other research has shown as many problems with search as with analysis, on this task students seem to have the greatest difficulty with synthesizing evidence.
2. Provide external representations of evidence that highlight where the mass of evidence contradicts belief. The study shows that students do not possess a clear picture of how the bulk of of evidence supports or contradicts a particular causal claim. The first step in tutoring must be to provide students with the skills and tools to recognize when evidence supports/contradicts a claim.
3. Tutor an *explicit evidence* epistemic rule. Even if students recognize that the majority of evidence provided contradicts their belief, the evidence may still not be enough to change their belief. To improve performance on this task, the tutor must emphasize that students are not to make recommendations based on belief, but on explicit evidence that they can *provide.* We might hope that over time, this will lead to beliefs that are more susceptible to evidence.

The results of this study and the previous diagramming studies [19, 25], provide the basis for a policy reasoning tutor: the diagram studies show the utility and feasibility of using causal diagrams to make policy claims explicit and to improve reasoning while this study identifies students' primary difficulties in the policy reasoning process. The greatest remaining challenge is to demonstrate tutoring of diagram construction.

Acknowledgements. The research reported here was supported in part by the Institute of Education Sciences, U.S. Department of Education, through Grant R305B040063 to CMU. The opinions expressed do not represent views of the Institute or the U.S. Department of Education.

References

[1] Westen, D. (2007). *The political brain: The role of emotion in deciding the fate of the nation*. New York: Public Affairs.

[2] Lakoff, G. (2002). *Moral politics: How liberals and conservatives think*. Chicago: University of Chicago Press.

[3] Carnegie Corporation of New York, & CIRCLE: The Center for Information & Research on Civic Learning & Engagement. (2003). *The civic mission of schools*. New York: Carnegie Corp. of New York.

[4] Kuhn, D. (1991). *The skills of argument*. New York: Cambridge University Press.

[5] Voss, J. F., Greene, T. R., Post, T. A., & Penner, B. C. (1983). Problem solving skill in the social sciences. In G. H. Bower (Eds.), *The psychology of learning and motivation: Advances in research and theory* (pp. 165-213). New York: Academic Press.

[6] Voss, J. F., Tyler, S. W., & Yengo, L. A. (1983). Individual differences in the solving of social science problems. In R. F. Dillion, & R. R. Schmeck (Eds.), *Individual differences in cognition* (pp. 205-32). New York: Academic Press.

[7] Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2(2), 175-220.

[8] MacCoun, R. J. (1998). Biases in the interpretation and use of research results. *Annual Reviews in Psychology*, 49(1), 259-87.

[9] Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108(3), 480-98.

[10] Molden, D. C., & Higgins, E. T. (2005). Motivated thinking. In K. J. Holyoak, & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 295-317). New York: Cambridge University Press.

[11] Schwenk, C. R. (1995). Strategic Decision Making. *Journal of Management*, 21(3), 471.

[12] Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, 37(11), 2098-109.

[13] Taber, C. S., & Lodge, M. (2006). Motivated Skepticism in the Evaluation of Political Beliefs. *American Journal of Political Science*, 50(3), 755-69.

[14] Redlawsk, D. P. (2002). Hot Cognition or Cool Consideration? Testing the Effects of Motivated Reasoning on Political Decision Making. *The Journal of Politics*, 64(4), 1021-44.

[15] Kuhn, D., Amsel, E., & O'Loughlin, M. (1988). *The development of scientific thinking skills*. San Diego: Academic Press.

[16] Koslowski, B. (1996). *Theory and evidence: the development of scientific reasoning*. Cambridge, MA: MIT Press.

[17] Chinn, C. A., & Brewer, W. F. (2001). Models of Data: A Theory of How People Evaluate Data. *Cognition and Instruction*, 19(3), 323-93.

[18] Zimmerman (2000). The development of scientific reasoning skills. *Developmental Review*, 20(1), 99-149.

[19] Easterday, M. W., Aleven, V., Scheines, R., & Carver, S. M. (in press). Constructing causal diagrams to learn deliberation. *International Journal of Artificial Intelligence in Education*.

[20] Edelson, D. C., & Reiser, B. J. (2006). Making authentic practices accessible to learners: Design challenges and strategies. In R. K. Sawyer (Ed.), *Cambridge Handbook of the Learning Sciences* (pp. 335-54). New York: Cambridge University Press.

[21] Spirtes, P., Glymour, C., & Scheines, R. (2000). *Causation, prediction, and search* (2nd ed.). Cambridge, MA: MIT Press.

[22] Tetrad (version 4.3.9) [Computer software]. (2008). Pittsburgh, PA: Laboratory for Symbolic and Educational Computing, Department of Philosophy, Carnegie Mellon University.

[23] Lodge, M., McGraw, K., & Stroh, P. (1989). An impression-driven model of candidate evaluation. *American Political Science Review*, 83(2), 399-419.

[24] Kim, S., Taber, C. S. and Lodge, M. (Oct. 10, 2008). A Computational Model of the Citizen as Motivated Reasoner: Modeling the Dynamics of the 2000 Presidential Election. Retrieved Jan, 12, 2009 from SSRN: http://ssrn.com/abstract=1282323.

[25] Easterday, M., Kanarek, J., & Harrell, M. (in press). Design Requirements of Argument Mapping Software for Teaching Deliberation. In T. Davies, & S. P. Gangadharan (Eds.), *Online Deliberation: Design, Research, and Practice (Center for the Study of Language and Information - Lecture Notes)*. Stanford, CA: Standford University Center for the Study of Language and Information.