



Production blocking in brainstorming arguments in online group debates and asynchronous threaded discussions

Allan Jeong¹ · Ming Ming Chiu²

Accepted: 7 October 2020 / Published online: 23 October 2020
© Association for Educational Communications and Technology 2020

Abstract

Online group debates hosted in asynchronous threaded discussions can facilitate critical thinking between discussants (and increase deeper understanding of complex problems) by eliminating the need for turn-taking while formulating and presenting premises to support and challenge claims. Yet to be determined is to what extent does the current number of posts (premises from teammates, premises from the opposing team, supportive replies to premises, and opposing replies to premises) induce production blocking to disrupt the generation of new premises. Statistical discourse analysis was conducted on 1554 postings from online debates produced by four cohorts of students enrolled in an online graduate-level course. The resulting stochastic model revealed that: (a) the number of premises posted by the other team, premises posted by teammates, and supportive replies from teammates were associated with the largest to smallest drop in the likelihood of posting a new premise, respectively; and (b) the number of oppositional replies from the other team was associated with an increased likelihood of posting a new premise. These findings provide evidence to support the use of specific strategies for structuring and sequencing the argumentation task to generate larger numbers of premises to achieve a deeper and more thorough analysis of problems and claims.

Keywords Online discussions · Collaborative argumentation · Discourse analysis · Computersupported · Collaborative learning

✉ Allan Jeong
ajeong@fsu.edu

¹ Instructional Systems and Learning Technology, College of Education, Florida State University, 3205E Stone Building, 1114 W. Call Street, Tallahassee, FL, USA

² Analytics and Diversity, Special Education and Counseling Department/Director of the Assessment Research Center, The Education University of Hong Kong, B1-2/F-15, 10 Lo Ping Road, Tai Po, New Territories, Ting Kok, Hong Kong

Introduction

Collaborative argumentation is an instructional activity that uses critical discussions and reflection to help students engage in deep thinking and learning—a process that involves integrating and applying multiple perspectives, prior knowledge, information, and personal experiences to thoroughly evaluate and understand complex problems (Andriesen and Coirier 1999; Chin and Brown 2000). Argumentation is the process of identifying and building arguments to support a position, considering and weighing evidence and counter-evidence (or grounds), presenting warrants to explain how the evidence or data support the claim, and presenting backing, rebuttals, and qualifiers (Toulmin 1958). As a result, argumentation can serve as a means to test uncertainties, extract meaning, achieve understanding, and examine complex ill-structured problems and issues (Jonassen and Kim 2010; Kuhn 1993). Formulating and generating reasons for and against conclusions and propositions is a fundamental part of human reasoning and decision-making (Mercier and Sperber 2011). Furthermore, improvements in idea generation lead to increased learning in scientific inquiry tasks (Wang et al. 2007). In other words, group performance and outcomes are dependent on and linked to the group's ability to generate ideas and arguments (Gouran 2000; Meyers et al. 2000). In collaborative argumentation, groups brainstorm and identify as many reasons as possible to support and oppose presented claims and propositions to conduct a thorough analysis and evaluation of each claim to make the best possible informed conclusions and decisions.

However, early studies show that interactive group brainstorming (when members take turns presenting ideas) produces fewer and less creative ideas than nominal groups where group members brainstorm alone before pooling and discussing their ideas (Connolly et al. 1990; Diehl and Stroebe 1987; Fjermestad and Hiltz 1999; Gallupe et al. 1992; Mullen et al. 1991; Nijstad and Stroebe 2006; Paulus et al. 1993). Interactive brainstorming produces fewer ideas because group members do not have sufficient time to think of new ideas, are distracted, and forget their ideas while attending to ideas presented by others in the group – conditions that contribute to production blocking (Diehl and Stroebe 1991; Nijstad et al. 2003; Sio et al. 2017). In a similar vein, group members tend to engage in collaborative fixation where members conform their ideas to the ideas of other group members while taking turns sharing ideas, which narrows the domains considered and used by the group to explore new ideas, which in turn contributes to a reduction in the number of novel ideas (Kohn and Smith 2011). Interactive brainstorming can also decrease idea generation because members fear negative evaluations of their ideas from other group members, particularly when issues under discussion are controversial (Camacho and Paulus 1995; Connolly et al. 1990; Dubrovsky et al. 1991) and because individuals can engage in free-riding by leaving the job of brainstorming ideas to other group members (Connolly et al. 1993; Valacich et al. 1992). All of these contributing factors help to explain why the dominant finding is that larger interactive groups generate no more ideas than smaller groups when working together face-to-face (Bouchard et al. 1974; Fern 1982; Lewis et al. 1975).

To counter production blocking and its negative effects on idea generation, studies have examined the effects of having group members work simultaneously on generating ideas (with no turn-taking). These studies (Connolly et al. 1993; Valacich et al. 1992) find that simultaneous idea generation enables groups to generate more ideas than nominal groups (the collective number of non-redundant individually generated ideas from all members in the group). At the same time, however, more ideas can be generated when working alone if the brainstorming process begins with simultaneous group brainstorming followed

immediately with individual brainstorming (Korde and Paulus 2017). When using networked computers in a computer laboratory to facilitate electronic brainstorming (EBS) and simultaneous idea generation in a shared group document, groups using EBS produce more and higher quality ideas than brainstorming in interactive groups and in nominal groups that use paper and pencil (Dennis and Valacich 1993; Fjermestad and Hiltz 1999; Gallupe et al. 1992; Nunamaker et al. 1987). However, Gallupe et al. (1991) found that the use of EBS with small groups, 4-person groups, did not produce more idea generation than nominal groups. As a result, subsequent studies (Valacich et al. 1994; Derosa et al. 2007) determined that EBS can help larger groups ranging from eight or more students (with increased opportunities for group synergy) generate more ideas than students working in smaller nominal groups.

The efficacy of group brainstorming on idea generation has also been found to be moderated by how the group task is presented and sequenced with studies showing that task sequencing (when tasks are broken down and performed in separate phases) can improve the number of generated ideas and depth of exploration over nominal groups (Baruah and Paulus 2011; Coskun et al. 2000; Deuja et al. 2014). Studies on the effects of task sequence on group performance show that the consideration of one topic or one category at a time (Coskun et al. 2000) and presenting subcategories of a broader problem (Rietzschel et al. 2007) help groups generate more and better ideas while sparking more in-depth discussion and exploration of ideas (Deuja et al. 2014). For example, Deuja et al. (2014) instructed students working individually or in three-person groups to post into AOL Instant Messenger all the possible ways in which the university can be improved. The study found that when the task/problem was broken down into categories in advance, brainstorming by category one at a time produced more ideas (but not so when the task was broken into categories by students themselves).

When students participate in online class debates (and in large teams of eight or more students per team) hosted in threaded discussion boards (with one team assigned to support and one team to oppose a given claim), the first task is to identify and post premises to support the team's position. Using threaded discussion boards, individual threads can be created in advance to break the debate down into separate tasks with separate threads designated for posting supporting premises and opposing premises (Brooks and Jeong 2006). The remaining tasks are to respond to the posted premises to provide supporting evidence and warrants, establish or challenge the veracity, relevance, and accuracy of each premise, and to respond to challenges with rebuttals to defend and justify each premise (Hemberger et al. 2017; Noroozi et al. 2013). In asynchronous online discussions, students can simultaneously and at any time read posted premises and post new premises to support their team's position. Most of all, students need not wait until all premises are posted before posting responses to justify a team member's premise and challenge an opposition's premise. Despite the expected benefits of using asynchronous threaded discussions to enable students to post ideas at any time, the quality of online class discussions, in general, are often shallow partly because students tend to focus attention and respond to new, flagged, and unread posts in online discussions. These behaviors can produce a starvation condition where specific discussion threads die from inactivity (Hewitt 2005) as well as divert students' attention away from the task of brainstorming and posting new premises to support their team's position. Studies that tested the efficacy of computer-supported argumentation systems – systems that provide scripts and prompts to guide students through each argumentation task – have largely been found to have little effect on group performance because the systems focus primarily on the structural components of argumentation while neglecting the social, discursive aspect of argumentation (Fischer et al. 2013; Noroozi et al. 2012).

As a result, the purpose of this study was to determine to what extent large groups exhibit production blocking when working simultaneously and asynchronously to brainstorm premises to support their team position while working without constraints on the task sequence and without the requirement that students finish posting all premise before discussing, challenging, and defending premises. More specifically, this study examined the extent to which the number and presence of particular types of postings (premises posted in support of the claim, premises posted in opposition to the claim, postings from supporting team members to strengthen the team's premises, and postings from the opposing team that challenges the team's premise) distract and hinder students' ability to identify and present new premises to support the team position in online debates hosted in an asynchronous threaded discussion board. At this time, no known studies have stochastically modeled and compared the extent to which different parts of the collaborative argumentation task (e.g., the current number of supporting premises posted, the current number of opposing premises posted) when performed simultaneously and in parallel (not sequentially across separate periods/phases) affects the generation of premises. As a result, no empirical research is available to inform online course instructors and designers on why, where, and when to feasibly sequence and structure online debates (and online discussions in general) to help increase idea generation and depth of discourse. To address this gap, this study examined the following four research questions. To what extent does the present number of:

- (1). *premises* posted by teammates affect the likelihood that teammates will post a new premise?
- (2). *opposing premises* posted by the other team affect the likelihood that teammates will post a new premise?
- (3). *supportive responses* from teammates that defend/justify the team's premises affect the likelihood that teammates will post a new premise?
- (4). *opposing responses* from the other team that challenges the team's premises affect the likelihood that teammates will post a new premise?

Method

Participants

The participants were four cohorts of graduate students ($N=87$, 53 females and 34 males, 20 to 50 years of age) enrolled in a 15-week online course titled Introduction to Distance Education at a major university in the Southeast region of the U.S. over a three-year span. Most students took all their courses at a distance in an online Master's program, and hence few of these students had opportunities to interact with one another outside of the course.

Procedure

In four of the 15 weeks of the online course, students participated in a week-long online team debate using asynchronous threaded discussion forums in the Blackboard™ web-based course management system. For each debate, students were randomly assigned to one of two teams (balanced by gender and level of participation across discussions from

previous weeks) to either support or oppose a given position. In each debate, students were required to post at least four messages per debate. At the end of each debate, students were required to vote on the team that presented the strongest arguments following each debate and to identify and explain their conclusions. Student participation in the debates and other weekly discussions and activities (e.g., discuss readings, curating research information, collaborative concept mapping, practice quizzes) scheduled over the 15 weeks contributed to 20% of the course grade.

The purpose of each debate was to critically examine design issues, concepts, and principles in distance learning covered during a particular week in the course. Students debated in weeks 3, 4, 6, and 7 of the semester over the following claims: “Type of media does not make any significant contribution to student learning”, “Synchronous chats should be used instead of asynchronous threaded discussion to host online team debates”, “Given the data and needs assessment, the fictitious country of NED should not develop a distance learning system”, and “Print is the preferred medium for delivering a course study guide”, respectively. In one of the four cohorts, the fourth debate topic was replaced with a debate over the claim “The instructional systems design model is an effective model for designing the instructional materials for this course” as a result of revisions to the course curriculum.

In Fig. 1 is a list of four message categories (with definitions and examples) based on Toulmin (1958) model of argumentation that was presented to the students before and during the debates to help them to identify premises labeled as ARG to either support or oppose the given claim (warrants), present evidence labeled as EVID to support a premise (facts), explanations labeled as EXPL (backing), and challenges labeled as BUT (rebuttals). Each student was required to classify each posted message by category by inserting the corresponding label into the subject headings of each message, and to restrict the content of their messages to address one and only one category at a time. Students were also instructed to identify each message by team membership by adding an “-” for opposing or a “+” for supporting team at the end of each label (e.g., +ARG, -ARG). Following each message tag posted in the subject heading, students were also instructed to insert a sentence to convey the main idea presented in the posting. These tagged subject headings allowed students to easily identify the ideas exchanged between the opposing and supporting teams during the debates (e.g., +ARG → BUT) and respond to the exchanges to advance their team’s position.

Students were instructed and reminded to read all premises posted by teammates and to read all responses to premises before posting a new premise to avoid redundancies in posted premises (typically one redundant post per thread). The course instructor occasionally checked students’ postings to determine if and to ask students to correct their message labels and to include a sentence following the label that conveys the main idea of the posting in the subject heading of their post. If requested by the instructor, students were allowed to return to a message to correct any errors in their labels. A student did not receive participation points for a given debate if the student failed to follow procedures.

Data preparation

The codes that were assigned to each message by the students were automatically pulled from the message subject headings to identify each message as either an argument (ARG), evidence (EVID), challenge (BUT), or explanation (EXPL). One debate from each course was randomly selected and coded by the investigator to test for errors in students’ message labels. The overall percent agreement was 0.91 based on the codes of 158 messages

Symbol	Description of symbol
+	Identifies a message posted by a student assigned to the team <u>supporting</u> the given claim/statement
-	Identifies a message posted by a student assigned to the team <u>opposing</u> the given claim/statement
ARG#	Identifies a message that presents <u>one and only one argument</u> for using or not using chats (instead of threaded discussion forums). Number each posted argument by counting the number of arguments already presented by your team. Sub-arguments need not be numbered. ARG = "argument".
EXPL	Identifies a reply/message that provides additional support, explanation, clarification, elaboration of an argument, or challenge.
BUT	Identifies a reply/message that questions or challenges the merits, logic, relevancy, validity, accuracy, or plausibility of a presented argument (ARG) or challenge (BUT).
EVID	Identifies a reply/message that provides proof or evidence to establish the validity of an argument or challenge.

Fig. 1 Screenshot of student instructions on how to label messages during the online debates

consisting of 42 arguments, 17 supporting evidence, 81 critiques, and 17 explanations. The Cohen Kappa coefficient, which accounts for the chance in coding errors based on the number of categories in the coding scheme, was 0.86—indicating excellent inter-rater reliability given that Kappa values of 0.40 to 0.60 is considered fair, 0.60 to 0.75 as good, and over 0.75 as excellent (Bakeman and Gottman 1997, p. 66).

Power analysis

As our data has multiple levels, we compute statistical power at each level. With 1554 messages, statistical power for even a tiny effect size of 0.1 at $\alpha \leq 0.05$ exceeds 0.97, but the four courses and 16 debates have little statistical power (< 0.80 even for a medium-size effect of 0.5, Konstantopoulos 2008).

Data analysis

Analyzing the above data in a manner that addresses a broad range of issues concerning the data, outcomes, and explanatory variables as described in Table 1 was achieved by using statistical discourse analysis (SDA) developed by Chiu and Lehmann-Willenbrock (2016). Regarding the data, an inadequate agreement among coders increases measurement error and *false negatives*. The Cohen's kappa *inter-rater reliability* as reported above was 0.86, above 0.80 to show excellent agreement (Blackman and Koval 2000).

Outcome issues include differences across messages or debates (nested data), similar adjacent messages (serial correlation), discrete outcomes, infrequency, and multiple outcomes. As messages within the same debate likely resemble one another more than those in different debates (*nested data*), an ordinary least squares regression underestimates the *standard errors*, so a *multilevel analysis* was used (Goldstein 2011; also known as *hierarchical linear modeling*, Raudenbush and Bryk 2002).

Failure to account for similarities in turns of talk within the same period or in adjacent turns (*serial correlation* of residuals) can underestimate the standard errors (Kennedy 2008). *Q-statistics* test all groups for serial correlation in adjacent turns (Ljung and Box 1979). If the serial correlation of the outcome (e.g., +ARG) is significant, adding the lagged outcome variable in the previous turn (+ARG [-1]) as an explanatory variable often removes the serial correlation (Jeong 2005).

For discrete outcomes (e.g., the message is +ARG vs. not), ordinary least squares regressions can bias the standard errors, so a *Logit regression* was used to model dichotomous outcomes correctly (Kennedy 2008). Infrequent events (occurring in less than 25% of the data) can bias logit regression results, so we estimate the bias and remove it (King and Zeng 2001). Multiple discrete outcomes can have correlated residuals that underestimate standard errors. This issue was addressed by using a multivariate outcome, multilevel model (Goldstein 2011).

Explanatory variable issues include sequences, indirect effects, interactions across levels, many hypotheses' false positives, comparison of effect sizes, and robustness. As preceding messages might influence the current message, the analysis must model previous sequences of message (Kennedy 2008). As a result, a *vector auto-regression* (VAR, Kennedy 2008) was used to test whether sequential characteristics of the recent message (*micro-time context*) influences the current message (e.g., the likelihood of +ARG).

Table 1 Addressing each analytic difficulty with statistical discourse analysis

Analytic difficulty	Statistical discourse analysis strategy
Data Set	
Inter-rater reliability	Cohen's kappa (Blackman and Koval 2000)
Dependent variables	
Differences across messages or debates (nested data)	Multilevel analysis (aka Hierarchical linear modeling, Goldstein 2011)
Similar adjacent talk ($t_3 \sim t_4$)	Q-statistics (Ljung and Box 1979)
Discrete variable (yes/no)	Logit/Probit (Kennedy 2008)
Infrequency	Logit bias estimator (King and Zeng 2001)
Multiple dependent variables (Y_1, Y_2, \dots)	Multivariate outcome model (Goldstein 2011)
Explanatory variables	
Sequences of turns of talk (X_{t-2} or $X_{t-1} \rightarrow Y_0$)	Vector Auto-Regression (VAR, Kennedy 2008)
Indirect, multi-level mediation effects ($X \rightarrow \mathbf{M} \rightarrow Y$)	Multilevel M- tests (MacKinnon et al. 2004)
Cross-level moderation (e.g., Student x Turn of Talk)	Random effects model (Goldstein 2011)
Many hypotheses' false positives	Two-stage linear step-up procedure (Benjamini et al. 2006)
Compare effect sizes ($\beta_1 > \beta_2?$)	Lagrange multiplier tests (Bertsekas 2014)
Consistency of results across data sets (Robustness)	Separate multilevel, single outcome models Analyses of subsets of the data (Kennedy 2008)

Separate, single-level tests of indirect mediation effects on nested data can bias results downward, so a *multilevel M-test* (MacKinnon et al. 2004) was used to test for simultaneous indirect multi-level effects. As the data are nested, modeling interaction effects across levels (e.g., Debate X Message) with a *fixed effects* model can bias the results (Goldstein 2011). This issue was addressed by using a *random effects* model (Goldstein 2011). If the regression coefficient of an explanatory variable (e.g., $\beta_{yvj} = \beta_{yv0} + f_{yvj}$) differs significantly across levels ($f_{yvj} \neq 0?$), then *cross-level moderation* might occur. As a result, structural variables (e.g., course) was used to model the regression coefficient.

As testing many hypotheses increases the possibility of a false positive, the likelihood of false positives was reduced by using the *two-stage linear step-up procedure*, which outperformed 13 other methods in computer simulations (Benjamini et al. 2006). When testing whether the effect sizes of explanatory variables differ, *Wald* and *likelihood ratio* tests do not apply at boundary points. Hence, *Lagrange multiplier tests* were conducted and applied to the entire data set, which resulted in greater statistical power than Wald or likelihood ratio tests for small deviations from the null hypothesis (Bertsekas 2014).

Lastly, two variations of the core model were used to test whether the results remain stable despite minor changes in the data or analysis (robustness, Kennedy 2008). Because a misspecified equation for any outcome in a multivariate outcome model can introduce errors in otherwise correctly specified equations, each outcome variable was modeled separately. Next, subsets of the data were run separately to test the consistency of the results across subsets.

Hypotheses tested

SDA was used to address answer each research question. In SDA, time constrains the direction of causality, so later processes cannot affect earlier attributes or processes. Course ID was entered first into the SDA (Chiu and Lehmann-Willenbrock 2016), followed by student demographics, and then attributes of student messages.

$$P(\mathbf{Premise}_{yij}) = F(\beta_y + f_{yj}) + e_{yij} \quad (1)$$

The probability, $P(\mathbf{Premise}_{yij})$, that the outcome y (supportive premise or critical premise) occurs at message i in debate j is its expected value via the Logit or Probit link function (F) of the overall mean β_y , and the debate-level unexplained component (*residuals*, f_{yj}).

$$P(\mathbf{Premise}_{yij}) = F(\beta_y + f_{yj}) + \beta_{yn}\mathbf{Course}_{yn} + \beta_{yuj}\mathbf{Current_Message}_{yij} + \beta_{yvj}\mathbf{Previous_Message}_{y(i-1)j} + \beta_{vj}\mathbf{Earlier_Message}_{y(i-2)j} + e_{yij} \quad (2)$$

Next, the explanatory variables were added. A vector of dichotomous variables controlled for each course (**Course**). As *likelihood ratio tests* are not reliable for multilevel analysis of binary outcomes, *Wald tests* were used to identify significant links (Goldstein 2011). Because only omitting *significant* variables might yield *omitted variable bias*, non-significant variables were removed from the analysis to reduce multicollinearity (Kennedy 2008).

Then, **Current_Message** variables were added: past premises by teammates, past premises by opponents, past teammate replies to team premises, past opponent replies to team premises, past team replies to opponent premises, past opponent replies to opponent premises. Each u regression coefficient ($\beta_{yuj} = \beta_{yu} + f_{yj}$) was tested for different links across threads and debates ($f_{yj} \neq 0$? *random effects* model, Goldstein 2011).

A significance level of 0.05 was used. Q-statistics tested all groups simultaneously for serial correlation, which was modeled when necessary (Ljung and Box 1979). The two-stage linear step-up procedure controlled for false positives (Benjamini et al. 2006). To aid in understanding the impact of each finding/result, the odds ratio of each variable's total effect (E ; direct plus indirect) was computed and reported as the percentage increase or decrease ($+E\%$ or $-E\%$) in the likelihoods of a *supportive premise* or *critical premise* (Kennedy 2008).

Data set and summary statistics

Table 2 shows how the 1554 messages posted to the 16 debates by the four student cohorts are distributed across the eight categories of postings (e.g., teammate premise, opponent premise). For example, supportive premises composed 8.6% of all postings whereas oppositional premises composed of 7.3% of all postings. Students in the first course wrote more messages (30.8%) than those in the fourth course (19%). Table 3 shows the descriptive statistics for the main variables. For example, the mean depth level (the posting within a thread with the maximum number of indentations) across all postings was 2.75 with a maximum thread depth of 9. For any postings made to a debate, the average number of premises already posted by teammates was 6.98 versus a mean of 7.05 premises already posted by the opponents.

Table 2 Proportion of postings by category

Postings	Mean	SD	Min	Median	Max
Course1	0.308	0.462	0	0	1
Course2	0.294	0.456	0	0	1
Course3	0.208	0.406	0	0	1
Course4	0.19	0.393	0	0	1
Teammate premise	0.086	0.281	0	0	1
Teammate challenge	0.232	0.422	0	0	1
Teammate explanation	0.086	0.281	0	0	1
Teammate evidence	0.083	0.275	0	0	1
Opponent premise	0.073	0.26	0	0	1
Opponent challenge	0.275	0.447	0	0	1
Opponent explanation	0.082	0.274	0	0	1
Opponent evidence	0.083	0.276	0	0	1

N = 1554

Table 3 Summary statistics of the main variables

Variable	Mean	SD	Min	Median	Max
Depth	2.751	1.396	1	2	9
Past premises by teammates	6.982	2.891	0	7	16
Past premises by opponents	7.057	3.091	0	7	16
Past teammate replies to team premises	11.158	7.917	0	9	39
Past opponent replies to team premises	9.278	7.82	0	8	46
Past team replies to opponent premises	9.833	8.133	0	8	45
Past opponent replies to opponent premises	10.666	7.712	0	9	39

N = 1554

Explanatory model

Neither type of premise differed substantially across debates. Whether a supportive premise occurred or not mostly differed across messages (100%) rather than across debates (0%, see Table 4). Likewise, whether a critical premise occurred or not mostly differed across messages (99%) rather than across debates (1%).

In this study, each debate forum presented one discussion thread designated for posting all premises in support of the given claim. Positioned below this thread was a second and separate thread for posting all premises in opposition to the claim. As a result, all supporting premises were visually presented in the upper half of the forum, and all opposing premises were presented in the lower half of the forum. As the number of posts in the first thread for posting and discussing supporting premises grew in number, the second thread containing the opposing premises were pushed progressively lower down the discussion forum. Once the first thread for posting and discussing supporting premises reached 30 or more postings, the second thread containing the opposing premises were bumped and scrolled off the page and out of immediate view when students open the debate forum. At

that point, students assigned to the supporting team to post supporting premises did not necessarily see the opposing premises displayed in the lower half of the debate forum. In contrast, students assigned to the opposing team by default saw all the supporting premises when they opened the debate forum and scrolled down the forum to access the thread for viewing and posting opposing premises.

As a result, the statistical discourse analysis was performed in two parts – one part to produce findings on the likelihood of posting a premise in *support* of the claim in relation to the presence and the current number of supporting premises, opposing premises, supportive replies posted to defend a supporting premise, and opposing replies posted to challenge a supporting premise. The second part of the analysis produced findings on the likelihood of posting a premise in *opposition* to the claim in relation to the presence and the current number of supporting premises, opposing premises, supportive replies posted to defend an opposing premise, and opposing replies posted to challenge an opposing premise.

Main findings

Posting supporting premises to the supporting premise thread

With the addition of each new supporting premise, new opposing premise, new supportive reply, and new opposing reply posted to the debates, the probability that the next posting from a student on the supporting team will be a new supporting premise dropped by $-5%$, $-6%$, $-3%$, and increased by $+3%$, respectively. Note that these effects are cumulative though non-linear. For example, the net effect of two opposing premises on the probability of a supportive premise in the next post, assuming independence, is $-11.6\% = -6\% + -6\% + -6\% * -6\%$. As a result, the findings suggest that the number of supporting premises, opposing premises, and supportive replies to supporting premises all contribute to production blocking and hinders the process of identifying new premises to support the team position. In contrast, the presence of opposing replies to presented supporting premises increased the likelihood of posting a new supporting premise. All four observed probabilities were statistically significant at $p < 0.05$ and were computed using the odds ratios from the regression coefficients (e.g., $-5\% = \text{odds ratio } (-0.098)$ from Table 3, supportive premise, model 2). See Brooks and Jeong (2005, p. 613) for more details on how the odds ratios are computed. The final model (model 2) accounted for an estimated 8.3% of the variance in supportive premise and 9.4% of the variance in opposing premise (see Table 3, bottom).

The first three of these four findings suggest that the presence of *supporting premises* and the presence of *opposing premises* contributed 1.66 and two times, respectively, more to production blocking than the presences of *supportive replies*. The observed impact of the presence of supporting premises can be attributed not only to the effects of diverting attention away from the process of identifying new supporting premises but also to the process of diminishing returns. In other words, the observed impact of supportive premises on production blocking (-5%) is likely to be inflated because identifying a unique premise to add to an existing list of premises becomes increasingly more difficult as all the possible premises (particularly the low hanging fruit) are fleshed out earlier in the debate week. On the other hand, the effects of diminishing return does not apply when examining the impact

Table 4 Summary of two models of multivariate outcome, multilevel logit regressions with unstandardized regression coefficient (and standard errors in parentheses)

Explanatory variables	Supportive premise		Opposing premise	
	Model 1	Model 2	Model 1	Model 2
# of prior supportive premises	-0.095* (0.038)	-0.098 (0.038)	-0.126 (0.039)	-0.128*** (0.039)
# of prior opposing premises	-0.124*** (0.037)	-0.111*** (0.036)		
# of prior supportive replies by teammates	-0.061** (0.019)	-0.055*** (0.019)		
# of prior opposing replies from opposing team	0.053** (0.017)	0.055** (0.017)	-0.090*** (0.022)	-0.098*** (0.022)
Oppositional challenges (lag 1)		-0.692*** (0.268)		
Supportive premises (lag 1)		0.656*** (0.235)		
Opposing explanations (lag 1)				0.610* (0.284)
Debate	Variance 0% 0.000	Explained 0.000	Variance 1% 0.000	Explained 0.000
Message	100% 0.064	0.083	99% 0.091	0.095
Total variance explained	0.064	0.083	0.090	0.094

of *opposing* premises because opposing premises were conceptually orthogonal to all supporting premises. As a result, the measured impact of supportive premises on production blocking should hypothetically be greater than the impact of opposing premises on production blocking. However, the results reveal the opposite trend (though small in difference) with opposing premises affecting production blocking (-6%) more than supporting premises (-5%). These findings overall suggest that the impact of opposing premises on production blocking may need to be given greater consideration than the impact of supportive premises and the impact of supportive replies.

In contrast, the fourth finding showed that the presence of *opposing replies* was associated with a greater likelihood (not lesser likelihood) of posting a supporting premise ($+3\%$). As a result, this finding suggests that the presence of opposing replies does not result in production blocking, but instead, encourages the production of supporting premises. With the average number of opposing replies to supporting premises at 9.27 at any point in time within a debate, the presence of opposing replies helped to spark the overall average production of roughly 1.38 supporting premises per debate and a 3% greater likelihood of posting a supporting premise each time a student on the supporting team visits the debate forum to make a posting. This 3% greater likelihood corresponds to a 99% greater likelihood within each debate across all students posting to the debate—computed from the mathematics unions of the marginal effect of each variable based on odds ratios (Kennedy 2008). One plausible explanation for this finding is that the students on the supporting team became more intent on adding more supporting premises as earlier premises were refuted by students on the opposing team.

Posting new opposing premises in the opposing premise thread

The number of opposing premises was linked to the likelihood of posting a new premise in the opposing premise discussion thread. Adding an opposing premise above the mean was linked to a lower likelihood (-6%) of posting a new opposing premise. This finding corroborates the finding of production blocking observed in the supporting premise thread where there was a decreased likelihood of posting new supporting premises (-5%) with an additional supporting premise above the mean. The number of supportive premises posted by the other team was not significantly linked to the likelihood of posting a new opposing premise. This finding did not corroborate the findings observed in the supporting premise thread where the likelihood of posting a new supporting premise decreased at -6% for an additional opposing premise above the mean.

The number of replies posted to challenge an opposing premise was linked to the likelihood of posting a new opposing premise, decreasing by -5% for an extra reply above the mean posted to challenge an opposing premise in the opposing premise discussion thread. This finding, however, conflicts with the findings observed in the supporting premise thread where the likelihood post posting a new supporting premise increased by $+3\%$ (not decreased) with an additional reply above the mean posted to challenge a supporting premise. Finally, the link between the number of replies posted to support a premise from teammates was not significantly linked to the likelihood of posting a new opposing premise. This finding did not corroborate the finding of production blocking observed in the supporting premise thread where the likelihood of posting a new supporting premise decreased by -3% with an extra reply above the mean posted to support a supporting premise.

Overall, the one finding from the analysis of postings to the opposing premise discussion thread that corroborates the findings from the supporting premise thread was the

observed link between the number of premises posted by the other team and the likelihood of posting and adding a new premise for the team. This corroboration lends support to the conclusion that the number of premises posted by the other team can lead to production blocking and a decrease in the likelihood of posting a new premise by anywhere from -5 to -6% . As for the inconsistencies observed in the other three findings between the supporting versus opposing premise threads, the numbers do not reveal any notable patterns and hence the reasons for the discrepancies are not immediately evident. Nevertheless, the observed inconsistencies could be the result of the way the opposing premises were displayed in the lower half of the forum each time students logged into the debate forum, and how students had to scroll through all the supporting premises (including replies posted to challenge and support the premises) to access the opposing premises. The observed inconsistencies between these two sets of findings serve to illustrate how idiosyncrasies in a discussion forum interface can affect user behaviors and idea generation in online discussions.

Instructional implications

Overall, this study provides some findings to support the idea of sequencing three particular tasks of argumentation (identify supporting premises, identifying opposing premises, support premises, but not challenge premises) to maximize the number of generated premises in support of the team's position. Based on the analysis of corroborating findings and a comparison of the observed likelihoods of posting new premises, the specific task that appears to induce the most production blocking and warrants the most attention is the posting and viewing of premises *posted by the other team*. One solution is set the discussion forum to open in collapsed view so that only the parent messages (the thread for posting supporting premise, and the thread for posting opposing premises) are displayed on the screen initially, and so that students must click on a thread to expand, view, and respond to the premises posted within each thread. Another solution is to create separate discussion forums to host supporting premises and to host opposing premises. A third solution is to instruct students to work together with teammates to post new premises and then immediately follow that by working individually to identify additional premises, as suggested by Korde and Paulus (2017) findings that show how alternating group and individual brainstorming can produce more ideas than working in groups or working individually alone. A fourth solution is to instruct students to refrain from reading and posting replies to the other team's premises until after a particular day in the debate week.

If the findings from the analysis of the opposing premise thread were to be dismissed due to the idiosyncrasies of the forum interface (as explained above), other instructional implications can also be drawn from the observed association between production blocking and the number of premises *posted by teammates* (but of lesser concern than blocking caused by premises posted by the other team as discussed earlier). In addition to the solutions presented above, one way to reduce production blocking caused by the number of premises posted by teammates is to create two separate forums for posting supporting and opposing premises, configure each forum to show earlier postings only *after* a student submits a premise to the forum (a feature that is available in Blackboard discussions), and instruct students to immediately review prior premises posted to the forum to determine whether or not to delete the just-posted premise if found to be redundant to a premise already posted in the forum. However, students would likely find this process cumbersome and inefficient. As a result, the alternative method of starting with group brainstorming

and then alternating to individual brainstorming as described above seems to be the more feasible solution.

Finally, the production blocking associated with *supportive replies from teammates* to support a team's premise (but at only half the level of impact as the number of premises posted by the other team and number of premises posted by teammates) can be resolved by dividing the debate into phases so that only premises can be posted during the first three days or after, and then after day three, students can post supportive replies to strengthen the premises posted by teammates. The finding that replies from the other team that challenged the premises helped to increase (not decrease) the likelihood of posting a new premise suggests that more encouragement can be given to students to challenge the premises to further motivate students to think harder to generate additional premises to make up for the loss of premises debunked by the other team. The finding also suggests that students should be allowed to challenge the premises of the other team at any time from the very first day of the debate (not just on and after day 4, for example).

Directions for future research

Overall, this study presents a preliminary stochastic model that reveals some of the discourse processes in large discussion groups that induce as well as counter production blocking in idea generation in online discussions – a step that other researchers believe is necessary to advance the design and implementation of computer-supported collaborative argumentation that has thus far produced little impact on group performance. Towards this goal, the resulting model provides preliminary evidence to help determine why, where, and when to use existing strategies to structure online discussions in ways that facilitate idea generation. Some of the limitations of this study and the suggested instructional strategies provide possible directions for further research.

To build on the findings produced in this study, future studies can: (1) conduct controlled experiments using inferential statistics combined with SDA to test the numerous suggested strategies and solutions to verify the cause effect relationships between production blocking and type of postings found to be associated with production blocking; (2) conduct eye gaze analysis combined with talk-aloud protocol to verify and directly examine the presence of some of the factors (e.g., diverted attention, fear of negative evaluation, free riding) that were assumed to contribute to production blocking in this study; (3) implement a protocol to immediately remove new premises found to be redundant before the posted premise elicits replies from other students to achieve more accurate analysis of discourse processes that induce production blocking; (4) test the generalizability of this study's findings by analyzing debates without using message labels – labels the enabled students in this study to easily distinguish and recognize supportive replies from teammates from the oppositional replies posted by the other team; (5) measure the diversity of posted premises by examining the number of major and minor premises as two separate outcome variables; and (6) conduct SDA on debates where students collaboratively create an argument diagram to help students explicitly distinguish major from minor premises and identify redundancies in premises.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Andriesen, J., & Coirier, P. (1999). *Foundations of argumentative text processing*. Amsterdam: Amsterdam University Press.
- Bakeman, R., & Gottman, J. M. (1997). *Observing interaction: An introduction to sequential analysis*. Cambridge University Press.
- Baruah, J., & Paulus, P. B. (2011). Category assignment and relatedness in the group ideation process. *Journal of Experimental Social Psychology, 47*, 1070–1077.
- Benjamini, Y., Krieger, A. M., & Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika, 93*, 491–507.
- Bertsekas, D. P. (2014). *Constrained optimization and Lagrange multiplier methods*. New York: Academic.
- Blackman, N. J. M., & Koval, J. J. (2000). Interval estimation for Cohen's kappa as a measure of agreement. *Statistics in Medicine, 19*(5), 723–741.
- Bouchard, T., Barsaloux, J., & Drauden, G. (1974). Brainstorming procedure, group size, and sex as determinants of the problem-solving effectiveness of groups and individuals. *Journal of Applied Psychology, 59*(2), 135.
- Brooks, D., & Jeong, A. (2006). The effects of pre-structuring discussion threads on group interaction and group performance in computer-supported collaborative argumentation. *Distance Education, 27*(3), 371–390.
- Camacho, L. M., & Paulus, P. B. (1995). The role of social anxiousness in group brainstorming. *Journal of personality and social psychology, 68*(6), 1071–1080.
- Chin, C., & Brown, D. E. (2000). Learning in science: A comparison of deep and surface approaches. *Journal of Research in Science Teaching, 37*(2), 109–138.
- Chiu, M. M., & Lehmann-Willenbrock, N. (2016). Statistical discourse analysis. *Group Dynamics: Theory, Research, and Practice, 20*(3), 242–258.
- Connolly, T., Routhieaux, R. L., & Schneider, S. K. (1993). On the effectiveness of group brainstorming: Test of one underlying cognitive mechanism. *Small Group Research, 24*(4), 490–503.
- Connolly, T., Jessup, L., & Valacich, J. (1990). Effects of anonymity and evaluative tone on idea generation in computer-mediated groups. *Management Science, 36*, 689–703.
- Coskun, H., Paulus, P. B., Brown, V., & Sherwood, J. J. (2000). Cognitive stimulation and problem presentation in idea generation groups. *Group Dynamics: Theory, Research, and Practice, 4*, 307–329.
- Dennis, A. R., & Valacich, J. S. (1993). Computer Brainstorms: More heads are better than one. *Journal of Applied Psychology, 78*, 531–537.
- Derosa, D. M., Smith, C. L., & Hantula, D. A. (2007). The medium matters: Mining the long-promised merit of group interaction in creative idea generation tasks in a meta-analysis of the electronic group brainstorming literature. *Computers in Human Behavior, 23*(3), 1549–1581.
- Deuja, A., Kohn, N., Paulus, P., & Korde, R. (2014). Taking a broad perspective before brainstorming. *Group Dynamics: Theory, Research, and Practice, 18*, 222–236.
- Diehl, M., & Stroebe, W. (1987). Productivity loss in brainstorming groups: Toward the solution of a riddle. *Journal of Personality and Social Psychology, 53*, 497–509.
- Diehl, M., & Stroebe, W. (1991). Productivity loss in idea-generating groups: Tracking down the blocking effect. *Interpersonal Relations and Group Processes, 61*(3), 392–403.
- Dubrovsky, V. J., Kiesler, S., & Sethna, B. N. (1991). The equalization phenomenon: Status effects in computer-mediated and face-to-face decision-making groups. *Human-computer Interaction, 6*(2), 119–146.
- Fern, E. (1982). The use of focus groups for idea generation: The effects of group size, acquaintanceship, and moderator on response quantity and quality. *Journal of Marketing Research, 19*(1), 1–13.
- Fischer, F., Kollar, I., Stegmann, K., & Wecker, C. (2013). Toward a script theory of guidance in computer-supported collaborative learning. *Educational Psychologist, 48*(1), 56–66.
- Fjermestad, J., & Hiltz, S. R. (1999). An assessment of group support systems experimental research: methodology and results. *Journal of Management Information Systems, 15*, 7–149.
- Gallupe, R., Bastianutti, L., & Cooper, W. (1991). Unblocking brainstorming. *Journal of Applied Psychology, 76*(1), 137.

- Gallupe, R., Dennis, A., Cooper, W., Valacich, J., Bastianutti, L., & Nunamaker, J. F. (1992). Electronic brainstorming and group size. *Academy of Management Journal*, *35*, 350–369.
- Goldstein, H. (2011). *Multilevel statistical models*. Sydney: Edward Arnold.
- Gouran, D. S. (2000). The theoretical adequacy of the argument-centered perspective on decision-making and problem-solving in groups. In T. A. Hollihan (Ed.), *Argument at century's end: Reflecting on the past and envisioning the future* (pp. 40–49). Annandale, VA: National Communication Association.
- Hemberger, L., Kuhn, D., Matos, F., & Shi, Y. (2017). A dialogic path to evidence-based argumentative writing. *The Journal of the Learning Sciences*, *26*(4), 575–607.
- Hewitt, J. (2005). Toward an understanding of how threads die in asynchronous computer conferences. *The Journal of the Learning Sciences*, *14*(4), 567–589.
- Jeong, A. C. (2005). A guide to analyzing message-response sequences and group interaction patterns in computer-mediated communication. *Distance Education*, *26*(3), 367–383. Retrieved October 1, 2020, from https://myweb.fsu.edu/ajeong/pubs/JJeong2005_AnalyzingInteraction.pdf.
- Jonassen, D. H., & Kim, B. (2010). Arguing to learn and learning to argue: Design justifications and guidelines. *Educational Technology Research and Development*, *58*(4), 439–457.
- Kennedy, P. (2008). *Guide to econometrics*. New York: Wiley-Blackwell.
- King, G., & Zeng, L. (2001). Logistic regression in rare events data. *Political Analysis*, *9*, 137–163.
- Kohn, N. W., & Smith, S. M. (2011). Collaborative fixation: Effects of others' ideas on brainstorming. *Applied Cognitive Psychology*, *25*(3), 359–371.
- Konstantopoulos, S. (2008). The power of the test in three-level cluster randomized designs. *Journal of Research on Educational Effectiveness*, *1*, 66–88.
- Korde, R., & Paulus, P. B. (2017). Alternating individual and group idea generation: Finding the elusive synergy. *Journal of Experimental Social Psychology*, *70*, 177–190.
- Kuhn, D. (1993). Science as argument: Implications for teaching and learning scientific thinking. *Science Education*, *77*(3), 319–337.
- Lewis, A., Sadosky, T., & Connolly, T. (1975). The effectiveness of group brainstorming in engineering problem solving. *IEEE Transactions on Engineering Management*, *3*, 119–124.
- Ljung, G., & Box, G. (1979). On a measure of lack of fit in time series models. *Biometrika*, *66*, 265–270.
- MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research*, *39*, 99–128.
- Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, *34*(2), 57–74.
- Meyers, R. A., Brashers, D. E., & Hanner, J. (2000). Majority-minority influence: Identifying argumentative patterns and predicting argument-outcome links. *Journal of Communication*, *50*(4), 3–30.
- Mullen, B., Johnson, C., & Salas, E. (1991). Productivity loss in brainstorming groups: A meta-analytic integration". *Basic and Applied Social Psychology*, *12*(1), 3–23.
- Nijstad, B. A., Stroebe, W., & Lodewijkx, H. F. M. (2003). Production blocking and idea generation: Does blocking interfere with cognitive processes? *Journal of Experimental Social Psychology*, *39*, 531–554.
- Nijstad, B. A., & Stroebe, W. (2006). How the group affects the mind: a cognitive model of idea generation in groups. *Personality and Social Psychology Review*, *10*(3), 186–213.
- Noroozi, O., Weinberger, A., Biemans, H., Mulder, M., & Chizari, M. (2012). Argumentation-based computer supported collaborative learning (ABCSCCL): A synthesis of 15 years of research. *Educational Research Review*, *7*(2), 79–106.
- Noroozi, O., Weinberger, A., Biemans, H., Mulder, M., & Chizari, M. (2013). Facilitating argumentative knowledge construction through a transactive discussion script in CSCL. *Computers and Education*, *61*, 59–76.
- Nunamaker, J. F., Jr., Applegate, L., & Konsynski, B. (1987). Facilitating group creativity with GDSS. *Journal of Management Information Systems*, *3*, 5–19.
- Paulus, P., Dzindolet, M., Poletes, G., & Camacho, L. (1993). Perception of performance in group brainstorming: The illusion of group productivity. *Personality and Social Psychology Bulletin*, *19*, 78–79.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. (Vol. 1). London: Sage.
- Rietzschel, E. F., Nijstad, B. A., & Stroebe, W. (2007). Relative accessibility of domain specific knowledge and creativity: The effects of knowledge activation on the quantity and originality of generated ideas. *Journal of Experimental Social Psychology*, *43*, 933–946.
- Sio, U., Kotovsky, K., & Cagan, J. (2017). The facilitating role of task alternation on group idea generation. *Journal of Applied Research in Memory and Cognition*, *6*(4), 486–495.
- Toulmin, S. E. (1958). *The uses of argument*. Cambridge: University Press.
- Valacich, J., Dennis, A., & Connolly, T. (1994). Idea generation in computer-based groups. *Organizational Behavior and Human Decision Processes*, *57*, 448–467.

- Valacich, J., Jessup, L., Dennis, A., Nunamaker, J.F. Jr. (1992). A conceptual framework of anonymity in GSS. *Proceedings of the 26th Hawaii International Conference on Systems Science*, 4, 101–112.
- Wang, H-C., Rosé, C., Cue, Y., Chang, C-Y. Huang, C-C., Li, T-Y. (2007). Thinking hard together: the long and short of collaborative idea generation in scientific inquiry. In *Proceedings of Computer Supported Collaborative Learning*, New Brunswick, NJ.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Allan Jeong is Associate Professor of Instructional Systems & Learning Technology at Florida State University. His research is in the field of computer-mediated communication and computer-supported collaborative learning creating learning analytic tools to visualize and model online discourse (*Discussion Analysis Tool*) and problem-solving processes (*jMAP* for creating and analyzing causal diagram structures and construction processes).

Ming Ming Chiu is Chair Professor of Analytics and Diversity (Honor) in the Special Education and Counseling department and Director of the Assessment Research Center at The Education University of Hong Kong. He invented *statistical discourse analysis* (SDA), *multilevel diffusion analysis* (MDA), artificial intelligence expert system *Statistician*, and online detection of sexual predators.