Pro/Con Lists and their Use in Group Decision Support Systems for Reducing Groupthink

MICHAIL TSIKERDEKIS

Faculty of Informatics
Masaryk University
Botanicka 68a, Brno, 602 00, Czech Republic
tsikerdekis@gmail.com

Abstract. Groupthink can occur under certain conditions in all collaborative groups and online groups that use group decision support systems such as wikis which are of no exception. One of the measures for preventing groupthink within a group is ensuring that all alternatives are considered and are also evaluated in detail. Pro/con list interfaces may be more powerful than traditional textual collaborative interfaces in wikis but no literature so far has tested the differences between the two in terms of argument production and comprehension beyond laboratory conditions using real world data. This study explored the potential effects of the two interfaces on group performance by conducting a comparative mix-methods analysis between two popular websites. The production of arguments, the uniqueness of arguments for each article and also the comprehension of articles were measured. No statistically significant differences between the two interfaces were found for the production of arguments and comprehension. However, the pro/con list interface statistically produced more unique arguments compared to the textual collaborative interface. In addition, a couple of qualitative remarks uncovered some of the limitations of both interfaces.

Keywords: pro/con list, wiki, interface, software, GDSS

(Received January 21st, 2012 / Accepted June 28th, 2012)

1 Introduction

Groupthink is a condition that can be found in collaborating groups when individuals choose to conform in order to minimize conflict and reach a consensus without critical evaluation of all options [19, 17, 16, 29, 30, 40]. Since the emergence of social media software, there has been an extensive amount of research on Group Decision Support Systems (GDSSs), especially wikis, and the tools that need to be used as preventive measures in order to avoid groupthink behavior [29, 30, 19, 14, 15, 25]. However, to this day there has been no adequate evaluation of pro/con lists or for their potential role in preventing groupthink behavior.

There are many recommendations to be considered in order to reduce the chances for groupthink some of which promote impartial leadership, exploring all alternatives, assigning devil's advocates, and introducing outside experts in the group [17, 16, 40]. Every single choice for the software design of GDSSs has potential implications in the effectiveness of the decision-making process. Hence, a design decision for using pro/con lists or a textual collaborative environment becomes less arbitrary and more essential.

The goal of this study is to provide software engineers with architectural design recommendations based on empirical results on whether the choice of pro/con lists is a viable solution for decreasing the chances of groupthink behavior for GDSSs. This paper begins with an overview on groupthink research, GDSSs, and comparisons between pro/con lists and other interfaces followed by the methods applied in this study for evaluating pro/con lists against textual collaborative interfaces.

A comparison between two websites that use the two interfaces is described in detail followed by several statistical analyses to assess efficiency in argument production and comprehension. Finally, recommendations are made for software developers and designers based on the results.

2 Theoretical Background

2.1 Groupthink

Groupthink was originally defined as "A mode of thinking that people engage in when they are deeply involved in a cohesive in-group, when the members' strivings for unanimity override their motivation to realistically appraise alternative courses of action" [16]. Several studies revised the original model and aimed to find ways to prevent groupthink [14, 36, 25, 29, 19].

One example of groupthink is the case of Pearl Harbor, an event that is considered a prime example of groupthink. The U.S. navy and army not only ignored all warnings but produced reports to rationalize why an invasion would be unlikely. One of the reports said, "Even if the Japanese were foolhardy to send their carriers to attack the United States, we could certainly detect and destroy them in plenty of time" [17]. Many similar case studies of fiascoes such as the Challenger accident [31], France's 1940 World War II defeat [1], Columbia accident [8], and the Iran hostage crisis [42] have shown positive evidence that groupthink has been the cause of defective decisions in all of these cases. Given the magnitude of the effect that poor decisions have due to groupthink, the need for research that could eliminate it became something of a holy grail for scientists. It became evident that preventing groupthink was essential to avoid future catastrophes such as these.

Based on literature, one can identify many antecedents and symptoms of groupthink such as gross omissions in surveying objectives and/or alternatives, failure to examine the costs and risks of the preferred choice, poor information seeking, selective bias in processing information at hand, failure to reconsider originally rejected alternatives, and failure to work out detailed implementation, monitoring and contingency plans [15, 40, 18]. Based on the symptoms several recommendations for preventing groupthink have been developed. For example, each member of a group should obtain the role of a critical evaluator, leaders should not express an opinion when assigning a task to a group and that several independent groups should work on the same problem [17, 40]. One important recommendation describes the need for consideration of all alternatives. The word alternatives could describe options to be considered or even arguments on a specific debate topic. Especially for weighing a decision, ensuring that a group has a complete view of all pro and con arguments is essential for successful decision-making; missing out on crucial information or alternatives may result in faulty decisions and actions.

As information technologies penetrate every part of our social and work life, a day will come that critical decisions may be taken from within GDSSs and social media software. Hence, understanding the effects of software tools in group performance could so to say, make or break a decision-making process for an online group.

2.2 Pro/Con Lists Compared to Other Interfaces

Pro/con lists have been used by humans throughout history to weigh overall arguments about topics and make better decisions. Ben Franklin identified the potential for weighing decisions through personal deliberation, and is identified as the inventor of modern pro/con lists [22]. ConsiderIt, a web platform that employs pro/con lists in its design to facilitate discussion has shown promising results that favor pro/con lists [23, 22]. But much of the scientific interest has been shifted away from pro/con lists towards argument maps and comparisons of the two. This in turn created a void in literature for pro/con lists which as a feature on today's Internet is more frequently occurring. However, argument maps are worth mentioning here since they have similarities with pro/con lists but also important differences.

Argument maps are a form of visual representation for the structure of arguments in evaluating argument production and more specifically the breadth and depth of a topic [5]. Breadth is defined as the number of distinct individual arguments advanced for and against certain positions, while depth is described as how extensively people elaborate on, distinct individual arguments by also, providing counter-arguments, rebuttals and evidence. Argument maps compared to pro/con lists that have been modified to measure depth, seem to be more effective and have produced more arguments breadth-wise and depth-wise [32]. Argument maps also seem to be more efficient for memory tasks in learning environments as opposed to a text version for information representations [10]. On the other hand, when it comes to comprehension and assimilation of information, argument maps are weaker than their text version counterparts. In addition, it stands to reason that argument maps do not come off as natural as pro/con lists or text versions, and there is a higher learning curve for participants willing to use them.

Pro/con lists may be a better alternative to textual

collaborative interfaces for GDSSs but there is no scientific evidence to this day to assert this claim. While argument maps produce more arguments compared to pro/con lists this cannot help us determine what happens when comparing pro/con lists with textual representations of arguments. Even in studies where students produced more information in argumentative maps than pro/con lists, they only partly internalized the collectively constructed arguments to construct their own arguments [41]. It may seem intuitive that organizing information may be more beneficial for the group members but when it comes to comprehension, results have shown the opposite [10]. Therefore, there is a need to close this gap between conventional textual collaborative interfaces and pro/con lists.

Two measurements were employed to evaluate and compare the two interfaces. The first, is the ability of a design to produce distinct arguments, that is earlier defined as breadth while the second accounted for the levels of comprehension and memorization.

2.3 Group Decision Support Systems & Wikis

Today, GDSSs software covers a broad spectrum of tools. It is generally defined as a collaboration technology that is designed to support meetings and group work [7]. The broadness of the definition includes online and offline collaboration environments. Benefits for online collaboration include more precise communication, members are empowered to build on ideas of others, and a more objective evaluation of ideas etc. [34]. Unfortunately, there are also certain tradeoffs such as information overloads, more flaming and slower feedback. Additionally, online collaboration is not immune to the effects of groupthink which can occur when a confluence of perfect circumstances exist between the group and other environmental factors [4, 17, 40]. In the case of online collaboration, the word "environmental" is the software that the participants are using aside from their own personal surroundings and inner cognition.

While software engineers have no control over the private space of individuals, they have control over the software. This control is evident by the amount of diverse software for collaboration that exists today. The differences between facilitate.com, MeetingSphere, ThinkTank and Wikipedia are great and many, even though all aim to facilitate online group collaboration.

Wikis are of particular interest for this study. A Wiki is a set of linked web pages created gradually by a collaborating group of users [26] as well as the software used for managing these web pages [21, 45]. Students generally find wikis as good tools for project collaboration [3]. Wikis can be seen as multipurpose collaborative environments. For example, people can use wikis for educational purposes [9] as well as in corporate environments. In the latter, people report that their groups are sustainable and wikis help them enhance their reputation, make their work easier and help the organization improve its processes [27]. Wikis also possess an extreme adaptability towards a team's goal producing a great variety of wikis that are modified to serve the needs of their communities. Wikis can be used to create collaborative research papers, encyclopedias and even facilitate debates [46].

According to the above, wikis are an essential link in the chain of online collaboration. They have the ability to allow for an unlimited number of users to participate in a discussion over vast distances of space and time! Users can provide their own remarks and allow for an asynchronous discussion that could take place over the course of years after which an article is created. Users have an infinite amount of time to evaluate information and contribute their own ideas which they can do even under the umbrella of complete anonymity which affects online collaboration [19]. As such wikis present an opportunity for testing a software when users have an unlimited amount of time to come up with arguments and are not restricted by artificially created experimental conditions that could exist in laboratory conditions.

3 Method

This study investigated what happens between two wiki websites in relation to their argument production process and the efficiency of comprehension provided to their users.

3.1 Measuring Production of Arguments

In order to measure argument production, a comparative content analysis of two wiki websites that used the two interfaces (textual collaborative interface and pro/con list interface) was employed. The first website was Wikipedia (`http://en.wikipedia.org`), a highly popular encyclopedia wiki and Debatepedia (`http://debatepedia.idebate.org`), a debate wiki which uses pro/con lists to facilitate discussion for the arguments of various debates. The two cover a wide variety of topics, some of which overlap between the two websites. This is the target population of topics that were used for the comparative study.

Both websites have similar goals that can be found on their about pages. Both are aiming for verifiability; every argument or piece of information must have a

source. The only significant difference is the language that users are allowed to use when writing. Wikipedia insists on the use of neutral language, while Debatepedia does not. In addition, Wikipedia as an encyclopedia, aims to cover arguments about topics but also general information which usually is not necessary for debates. In order to establish a base line and look for the same relevant arguments for each topic in both pages, a rule was devised. All arguments that were coded in Wikipedia and Debatepedia have to answer to the base question that exists at the top of a debate in Debatepedia. As an example, the question in the case of the article about the "*One Child Policy in China*" is "*Is China's 'one child' policy sensible?*" Each argument between the two pages used in comparison had to satisfy the key question in some way, which in effect became the core debate question of the topic for both website articles. In addition, all pages must have had a history tracing back at least 3 months in order to allow enough time for arguments to be gathered and reported by the members of each website.

Topics were obtained using the random function that existed on Debatepedia. After obtaining the article from Debatepedia, an attempt was made to find the same topic on Wikipedia. Topics required to have the same title in order to cover the same content. A total of 15 topics were analyzed for each website; a total of 30 articles for both websites.

Arguments were coded using content analysis and more specifically hermeneutic coding aimed to capture the latent meaning for the arguments that exist in the text. The content analysis helped operationalize key components within the text, and also to develop basic categorical distinctions for all arguments [24]. The software that was used in order to achieve this was Atlas.ti. It allowed the coding of arguments in one wiki page and then the use of the same categories when coding was applied on the other wiki page. This increased internal validity and reduced the interpretive effect because of the inevitable human factor. The general attitude towards coding was to create more arguments than less, and therefore creating categories that provide an extensive view for each argument. After both wiki pages were coded, categories that were similar got merged in order to be left with only unique and distinct categories.

As described in a previous section of this paper, the overall arguments of a topic could be analyzed based on breadth and depth [5]. However, unlike laboratory experiments where students can use software that allows them to provide depth in their arguments –by producing counter-arguments, rebuttals, as well as evidence–, the wiki websites do not provide such depth for their arguments. In order to avoid any bias that may arise by the researcher trying to adapt depth in content that was not designed to have depth, the decision was made to measure just the breadth of the arguments. Put simply, the comparative analysis measured the amount of unique and distinct arguments between the two wiki websites. In addition, the study also measured how many arguments overlap between each wiki in order to establish the overall uniqueness of produced arguments for each wiki page.

A note that has to be made here is that even though the study aimed to use quantitative measures to establish the final result and answer its hypotheses, it did not ignore findings that rose from the qualitative analysis.

3.2 Measuring Readability of Text

Literature describing the differences between argument maps and conventional text seems to be in favor of argument maps when it comes to comprehension [2, 44, 43, 10]. Based on this, it is fair to expect that participants will score better in comprehension, when it comes to pro/con lists versus traditional textual collaborative interfaces.

Measuring comprehension can be achieved by numerous ways. One approach is the use of surveys aimed to measure comprehension but using a survey to establish comprehension is also limited by the willingness of the participants to follow the survey's instructions to the letter. This in turn could produce unwanted variance within the data.

However, surveys are not the only way for evaluating comprehension. An indirect measure that could be used to determine memorization, comprehension, readership, reading persistence and reading speed, is a readability formula. Higher readability of a text improves all of these factors [20]. There are several widely accepted readability formulas. The Flesh Reading Ease (FRE) [11] is used to evaluate legal rights for defendants among Great Britain [13] and measures in a scale of 1 to 100, with 100 being the best score possible. An updated version of FRE is Flesch-Kincaid [12] and is considered the most reliable estimate of required reading comprehension with a high level of consistency across writing samples [35]. Another formula measuring the grade for public education that is reliable with a standard error of 1.5 grades is SMOG [28]. There is also the option of combining all these measures of readability to determine comprehension when conducting research [39].

This study employed all of the above readability formulas in order to establish a comparative quantitative analysis between articles with similar topics that were

retrieved from Wikipedia and Debatepedia. The aim is similar to the content analysis that was described in section 3.1 and the same articles used there were analyzed for readability as well.

4 Results

4.1 Content analysis

The analysis involved articles from various topics and of lengths between the two websites. Some of the articles described highly controversial issues such as China's one child policy, or Israel's flotilla raid in Gaza, while other articles described concepts such as the year-round school and the use of cell phones while driving. For a complete list of the articles for both websites see table 1. All articles were chosen after careful evaluation of the latent meaning and content of both articles in order to evaluate that they elaborate on the same issue.

Table 1: Articles that were subdued to content analysis

Wikipedia Article	Debatepedia Article
African Union	African Union
AIG bonus payments controv.	AIG Bonuses
Mobile phones and driv. saf.	Banning cell phones in cars
Full body scanner	Full-body scanners at airports
Gambling	Gambling
Health insurance cooperative	Health insurance cooperatives
Instant replay in baseball	Instant replay in baseball
Manned mission to Mars	Manned mission to Mars
Merit pay	Merit pay for teachers
Needle-exchange programme	Needle exchanges
One-child policy	China "one child" policy
Gaza flotilla raid	Israeli raid on Gaza flotilla
Women in the military	Women in the military
Keystone Pipeline	Keystone XL US-Can. oil pip.
Year-round school	Year-round school

In order to establish intercoder reliability for the content analysis that followed, a random sample was taken from the articles that were used in the study. Two Wikipedia and two Debatepedia articles which accounted for approximately 13% of the full sample were examined by a second coder. The size of the subsample was in accordance with literature on content analysis [47, 33]. The coder was trained to follow the same coding procedure that was used in the full sample. Agreement between the researcher's evaluation and the second coder's was established based on the total number of arguments found on each article. Krippendorff's alpha showed a high level of agreement between researcher and coder, $\alpha = .938$. This constitutes as a highly acceptable result [33].

At the initial stage of the analysis, one of the primary observations between the articles of the two websites was the variance in length. One of the prime examples of large size difference was the case of year-round schooling. The Debatepedia article contained 3,834 words whereas the Wikipedia article consisted of just 454 words. In some articles the opposite was found. In the case for the articles of the African Union, the Debatepedia article had a text of 1,344 words while the Wikipedia article contained 7,749 words. Since the study aimed to focus on argument production, the size of documents can be irrelevant, especially when it comes to the amount of arguments if one considers that Wikipedia's goal is to gather arguments as well as general information. Nevertheless, the differences in the size of the articles are remarkable and during the qualitative analysis it became apparent that size can have an effect on the incremental development of an article.

To understand the last statement, the year-round school example is ideal. While the Wikipedia article has existed since 2006, the development of the article has remained slow and tenuous. The article has had several claims that can also be found in the Debatepedia article, but has completely missed arguments in several areas such as economic and vacation related arguments. At the time of the analysis, the Wikipedia article in its history had 886 different users that made contributions to the article with minor and major revisions. Yet, the article did not expand of more than 500 words and major categorical arguments were missed. The total arguments for Wikipedia were 8 whereas for the Debatepedia article the arguments were 44. While this is an extraordinary case of an undeveloped article for Wikipedia, more research is required to understand the processes involved that led to an article remain underdeveloped even though it existed for 5 years and had so many people contributing to it.

On the opposite side of the seesaw, the article for the African Union not only was more detailed for Wikipedia but it managed to surpass the Debatepedia article in the amount of arguments; 21-20 respectively. The contributors for this Wikipedia article were 914 and the article was developed since 2002. Although this may show a certain balance between the articles, conclusions are harder to draw when taking into account the common arguments between the two articles which were only 9. This means that approximately 50% of each article had unique arguments that were not present in the other article. Effectively, both took different directions and explored different aspects of the same topic.

Even though Debatepedia's pro/con list interface may intuitively seem flawless, this may not be the case at all times. In several cases arguments for specific

claims were totally missed by members. Just as with Wikipedia and the various thematic categories, Debatepedia divides the pro/con lists with claims. Claims can evaluate a topic from various perspectives. In the case of AIG bonuses Wikipedia produced 49 arguments versus 41 arguments that Debatepedia produced. In one of Debatepedia's claims that was related to the taxation of the money that was given as bonuses, 4 arguments (2 pro and 2 con) were given, but several others that were found in the relevant Wikipedia article were never reported.

Furthermore, it seems that Debatepedia's members have an upper limit in the amount of arguments per claim and an average expected number for each claim. Beyond that average expected number, claims rarely had additional arguments added to them. In order to evaluate the above, arguments were measured for each claim of each topic. The total number of claims from all topics combined were 102. The mean number of arguments for the pro side of claims was 2.78 ($SD = 1.70$), while for the con side of claims it was 2.65 ($SD = 1.77$). Based on these mean values, it is reasonable to expect that the community will probably consider a claim satisfied by arguments when the arguments on each side are 2 or 3. In fact, from the 102 claims only 30 had surpassed the 3 argument threshold for the pro side and 28 for the con side. While it may be the case that people genuinely have no more arguments to report, cases such as the tax claim from the AIG bonuses topic, and the numbers presented above, seem to indicate that pro/con lists may produce an upper limit for the number of arguments per claim.

4.2 Statistical Analysis for Total and Unique Arguments

After the initial content analysis which helped code the unique arguments per article as well as the amount of arguments that were common between pairs of articles, normality analysis tests were performed. Since the sample was small, the Shapiro-Wilk test was used in evaluating the normality of each variable. While data were normal for Debatepedia($W = 0.964, df = 15, p = .762$), results for Wikipedia's data suggested that they were not normal ($W = 0.863, df = 15, p = .027$).

As mentioned in the previous paragraph, alongside with the arguments for each page, the common arguments between pages were also measured. This was then adjusted as a percentage value for each website based on the arguments that were found on each website. The formula that produced the percentage of unique arguments for each article was, $\text{uniqueness} = (1 - \frac{\text{common arguments}}{\text{total arguments in the article}}) \times 100$. This was a second important measure that could indicate the effectiveness of an interface over another. Normality tests indicated that Wikipedia($W = .948, df = 15, p = .487$) and Debatepedia($W = .946, df = 15, p = .467$) data were normal. Descriptive statistics of the two above mentioned variables are provided on table 2 and 3.

Table 2: Descriptive Statistics for Total Arguments

Website	Mean	SD	Max	Min
Wikipedia	25.47	16.88	64	7
Debatepedia	32.27	10.90	53	15

Table 3: Descriptive Statistics for Unique Arguments

Website	Mean	SD	Max	Min
Wikipedia	41.72	24.99	78.1	0
Debatepedia	59.84	16.76	86.4	35

While the means of Debatepedia(and in effect pro/con list interfaces) were higher than Wikipedia's, the significance of the variance could only be determined with an analysis of variance. Mann-Whitney U test was used due to normality violations to assess the variance for the total arguments between the two interfaces, $U = 68.5, Z = -1.827, p = .068, r = .334$. The result was approaching significance but did not achieve it. However, there is a medium effect size that is reflective of the difference between the means on table 2.

On the other hand, statistically significant differences were found for the uniqueness between the two samples using an independent t-test, $t(28) = 2.332, p = .027, r = .403$. This is a clear indication that Debatepedia produced more unique arguments than Wikipedia. This can be seen as a clear indication that pro/con lists are superior to conventional textual collaborative interfaces. The above argument can also be seen in the descriptive statistics provided by table 3.

4.3 Statistical Analysis for Readability

As mentioned in section 3.2, one of the targets of this study was to determine the level of comprehension that both interfaces provide to their users. Since understanding all arguments within a text is key in order to avoid groupthink behavior, readability plays a critical role in the efficiency of both interfaces.

The readability of each article was measured based on the formulas that can be found on table 4. The normality of the data produced by the readability formu-

Table 4: Readability Formulas

Readabil.	Readability formula
SMOG	$1.0430\sqrt{30 \times \frac{\text{number of polysyllables}}{\text{number of sentences}}} + 3.1291$
FRE	$206.835 - 1.015(\frac{\text{total words}}{\text{total sentences}}) - 84.6(\frac{\text{total syllables}}{\text{total words}})$
F-K	$0.39(\frac{\text{total words}}{\text{total sentences}}) + 11.8(\frac{\text{total syllables}}{\text{total words}}) - 15.59$

las was evaluated using Shapiro-Wilk analysis. All data within each group had a normal distribution which satisfied the assumption required by the independent t-test analysis. Results for normality tests are provided on table 5.

Table 5: Shaphiro Wilk normality tests for all readability samples

Readab. Sample	Saphiro-Wilk	df	Sig.
FRE Debatepedia	0.966	15	.789
Flesch-Kincaid Debatepedia	0.920	15	.190
SMOG Debatepedia	0.920	15	.196
FRE Wikipedia	0.985	15	.992
Flesch-Kincaid Wikipedia	0.991	15	1.000
SMOG Wikipedia	0.953	15	.576

The readability varied between each formula and website. As an example, in the case of the SMOG readability formula, Wikipedia had the highest score compared to Debatepedia but it also had the lowest. This result exists also for the Flesch-Kincaid measure. In addition, for all formulas the highest and lowest values are located far apart from each other for both websites. Hence, while one article may be easily read by individuals, another may require more cognitive energy so that it can be understood. A detailed report with all the means, standard deviations, as well as maximum and minimum values is presented on table 6.

Table 6: Descriptive statistics for Readability

Readab. Sample	Mean	SD	Max	Min
FRE Debat.	50.19	6.631	60.02	37.83
FRE Wikip.	46.67	7.161	59.15	31.50
F-K Debat.	9.85	1.288	12.49	7.96
F-K Wikip.	10.19	1.490	13.04	7.15
SMOG Debat.	11.87	1.003	13.95	10.53
SMOG Wikip.	12.12	1.099	14.62	9.80

An independent-samples t-test was conducted for each one of the three readability formulas. There was no statistically significant difference between Wikipedia's and Debatepedia's readability based on the FRE formula ($t(28) = -1.398, p = .173, r = .255$), the Flesch-Kincaid formula ($t(28) = 0.666, p = .511, r = .124$), and the SMOG formula ($t(28) = 0.649, p = .522, r = .122$).

5 Discussion

The results from the above study have been enlightening and unexpected at the same time. It seems that there is no clear-cut answer about which interface is better for producing more arguments and providing someone with better chances of comprehension, in order to reduce the chances of groupthink occurring within a group. Establishing ways to reduce the factors that contribute to groupthink at least from an interface design perspective, is challenging but also beneficial. Ensuring that users have enough alternatives available for consideration [17] could be affected by individuals as well as the virtual environment in which a collaborative task takes place. There is no literature that has ever measured the differences between pro/con lists and conventional textual collaborative interfaces.

Another issue with the way that differences such as these have been addressed by literature so far is the way that studies were conducted; short term laboratory experiments. This study used long term collaborative projects with multiple users coming from a variety of countries and having any information they may need literally at the end of their fingertips. This provided the users of both websites Wikipedia and Debatepedia with the best case scenario for producing enough argumentative coverage for the topics in question, and also enough time to revise and improve the text which in turn should improve readability and comprehension.

According to the results, pro/con lists did not seem to produce significantly more arguments than textual collaborative interfaces. This result is of great importance considering that there are many proponents for not only pro/con lists but also argumentative maps among the scientific community. However, results from this study cannot support a claim that both interfaces are equally as effective. In fact, what seems to be more essential for argument production is the initial segmentation of an article into categories that relate to the topic. These can later develop to include most of the arguments.

On the other hand, it is obvious that when the uniqueness of arguments within an article was evaluated, pro/con lists had a clear advantage against conventional text collaboration. Debatepedia users were more exhaustive upon finding at the very least some arguments for each claim, whereas Wikipedia users in many cases never had certain claims available to begin with. This seems to be a problem in the way that an ar-

ticle is segmented into various claims or categories. If the claims or categories are not there, arguments would never be considered by the users.

Finally, in the case of readability and comprehension, three different formulas were used in order to increase internal validity. In all cases no significant differences were found between Wikipedia and Debatepedia. It seems that both websites provide the same level of readability to their users. Whether having a textual collaborative interface or a pro/con list does not seem to affect readability and in turn comprehension. On the other hand, there is a big variability between the high and low scores for each readability formula for both websites. One may end up reading an article that has a readability of level 7 grade, while another article may require someone to be older than a 13th grader.

Although extreme cases such as the ones described above are bound to exist in online communities, it is highly unlikely that it will occur in a more controlled community such as a brainstorming group for a corporation or a high school class. Due to more authoritative control by the manager or the teacher results will probably be similar to the means seen in this study and probably would not have deviated beyond the standard deviation.

5.1 Recommendations

The implications from the results of this study are significant to software engineers of online collaborative communities and especially users of wiki software. The choice between textual collaborative interface or a pro/con list interface is much more than just aesthetics. Even though the readability from both interfaces was found to be similar, the overall uniqueness of the articles was not. In addition, the superiority of pro/con lists in producing more unique arguments, seems to be overshadowed by the same problems that exist in both interfaces. The initial structure of an article seems to affect the way that it will evolve by its users. The creation of categories upon the article's creation may affect the later development of an article and the arguments that it will produce. Pro/con list have an advantage because of their nature revolving around claims to produce arguments, but users may skip reporting all arguments under a claim that has already two or three arguments.

Based on the results of this study, it is recommended for software designers and developers to use pro/con list interfaces for controversial topics and especially when dealing with wicked problems of all sorts [38, 6]. However, the verdict for textual collaborative interfaces is not out yet. Given that the argument production and readability of the two websites did not vary significantly, a combination of both interfaces may also work just as effectively.

5.2 Limitations

Even though an interface that can produce more arguments can potentially reduce the chances for groupthink, it is still only one preventive measure. Groupthink can occur in a group regardless of an interface if other issues are not properly addressed. A wiki can be accessed by multiple users from various cultural and academic backgrounds, and from all over the world. That is usually not the case with corporate GDSSs or a class project.

Moreover, measuring comprehension indirectly with readability while it is cost efficient compared to surveys, is still an indirect measure for comprehension. Readability formulas also have several weaknesses while being compared to direct usability testing such as ignoring between reader differences or ignoring the effects of the content [37]. In addition, although readability has shown an increase in comprehension, the generalization of this should come with caution depending on the context of a study. A study has shown that better comprehension of a text does not necessarily mean efficiency in considering and internalizing all arguments for a debate [35]. Finally, by having a comprehensive text, it does not ensure that people would not still fall victims of cognitive dissonance and other phenomena that affect perception and decision-making.

5.3 Future Research

More research is required to understand the development of wiki articles; especially for both interfaces. Articles seem to be underlined by certain processes that affect their evolution and production of arguments. Additional studies are needed to understand how the initial categorization of an article affects the later development. In addition, pro/con lists should also be examined further and evaluate if indeed there is a norm or an expectation by their users to limit the amount of arguments per claim beyond a certain number.

6 Final Thoughts

This study is just a small piece of the puzzle for eliminating groupthink which is a phenomenon controlled by multiple factors. It is certain that as online collaboration software becomes ever more popular in the virtual world, decisions made by software designers and engineers become critical for determining the collaborative efficiency for online groups. Whether software is being

adapted for a teacher's class or a corporate brainstorming group, studies such as these help us understand how people behave in virtual collaborative environments. As more and more minds that think alike connect to the online world and extend their lives from the real to the virtual, software engineers will have the unique role that only ancient architects had when the great wonders of the world were being built. Creating today's virtual wonder world built for collaboration, is not an easy task but nonetheless a groupthink free environment should still be a goal that future generations to come, may one day enjoy.

References

[1] Ahlstrom, D. and Wang, L. C. Groupthink and France's defeat in the 1940 campaign. *Journal of Management History*, 15(2):159–177, 2009.

[2] Butchart, S., Forster, D., Gold, I., Bigelow, J., Korb, K., Oppy, G., and Serrenti, A. Improving critical thinking using web based argument mapping exercises with automated feedback. *Australasian Journal of Educational Technology*, 25(2):268–291, 2009.

[3] Chao, J. Student Project Collaboration Using Wikis. In *20th Conference on Software Engineering Education Training CSEET07*, pages 255–261, Washington, DC, 2007. IEEE.

[4] Chen, C.-k., Tsai, C.-h., and Shu, K.-C. An Exploratory Study for Groupthink Research to Enhance Group Decision Quality. *Journal of Quality*, 16(2):137–152, 2009.

[5] Chinn, C. A. and Andersen, R. C. The Structure of Discussions inteded to Promote Reasoning. *Teachers College Record*, 100(2):315–368, 1998.

[6] Conklin, J. *Dialogue Mapping: Building Shared Understanding of Wicked Problems*. John Wiley & Sons, Inc, New York, NY, 2005.

[7] Dennis, A. R., George, J. F., Jessup, L. M., Nunamaker, J. F., and Vogel, D. R. Information technology to support electronic meetings. *MIS Quarterly*, 12(4):591–624, 1988.

[8] Dimitroff, R. D., Schmidt, L. A., and Bond, T. D. Organizational behavior and disaster: A study of conflict at NASA. *Project Management Journal*, 36(1):28–38, 2005.

[9] Duffy, P. and Bruns, A. The Use of Blogs, Wikis and RSS in Education: A Conversation of Possibilities. In *Proceedings Online Learning and Teaching Conference 2006*, number 2006, pages 31–38, Brisbane, 2006. Queensland University of Technology.

[10] Dwyer, C. P., Hogan, M. J., and Stewart, I. The evaluation of argument mapping as a learning tool: Comparing the effects of map reading versus text reading on comprehension and recall of arguments. *Thinking Skills and Creativity*, 5(1):16–22, Apr. 2010.

[11] Flesch, R. A new readability yardstick. *Journal of Applied Psychology*, 32(3):221–233, 1948.

[12] Flesch, R. Measuring the level of abstraction. *Journal of Applied Psychology*, 34(6):384–390, 1950.

[13] Gudjonsson, G. H. The âNotice to Detained Personsâ, PACE Codes, and reading ease. *Applied Cognitive Psychology*, 5(2):89–95, Mar. 1991.

[14] Hart, P. Preventing Groupthink Revisited: Evaluating and Reforming Groups in Government. *Organizational Behavior and Human Decision Processes*, 73(2/3):306–326, 1998.

[15] Henningsen, D. D., Henningsen, M. L. M., Eden, J., and Cruz, M. G. Examining the Symptoms of Groupthink and Retrospective Sensemaking. *Small Group Research*, 37(1):36–64, 2006.

[16] Janis, I. L. *Victims of groupthink: a psychological study of foreign-policy decisions and fiascoes*. Houghton, Mifflin, Boston, MA, 1972.

[17] Janis, I. L. *Groupthink: psychological studies of policy decisions and fiascoes*. Houghton Mifflin, Boston, MA, 1982.

[18] Janis, I. L. and Mann, L. *Decision making: A psychological analysis of conflict, choice, and commitment*. Free Press, New York, NY, 1977.

[19] Jessup, L. M., Connolly, T., and Galegher, J. The effects of anonymity on GDSS group process with an idea-generating task. *MIS Quarterly*, 14(3):313–321, 1990.

[20] Karmakar, S. and Zhu, Y. Visualizing text readability. In *IEEE International Conference on Data Mining and Intelligent Information Technology Applications(ICMiA 2010)*, pages 291 – 296, New York, NY, 2010. IEEE.

[21] Khosrowpour, M. *Innovative Technologies For Information Resources Management*. Idea Group Inc (IGI), Hershey, PA, 2008.

[22] Kriplean, T., Morgan, J., Freelon, D., Borning, A., and Bennett, L. Supporting reflective public thought with considerit. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, pages 265–274, New York, NY, 2012. ACM.

[23] Kriplean, T., Morgan, J. T., Freelon, D., Borning, A., and Bennett, L. ConsiderIt: improving structured public deliberation. In *Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems*, pages 1831–1836, New York, NY, 2011. ACM.

[24] Krippendorff, K. *Content Analysis: An Introduction to Its Methodology*. Sage, 2004.

[25] Kroon, M. B. R., Hart, P., and Van Kreveld, D. Managing Group Decision Making Processes: Individual Versus Collective Accountability and Groupthink. *International Journal of Conflict Management*, 2(2):91–115, 1991.

[26] Leuf, B. and Cunningham, W. *The Wiki way: quick collaboration on the Web*. Addison-Wesley, Boston, MA, 2001.

[27] Majchrzak, A., Wagner, C., and Yates, D. Corporate wiki users. In Riehle, D. and Noble, J., editors, *Proceedings of the 2006 international symposium on Wikis - WikiSym '06*, volume pp of *WikiSym 2006 - The International Symposium on Wikis*, page 99, New York, NY, 2006. ACM, ACM Press.

[28] McLaughlin, G. H. SMOG grading: A new readability formula. *Journal Of Reading*, 12(8):639–646, 1969.

[29] Miranda, S. M. Avoidance of Groupthink: Meeting Management Using Group Support Systems. *Small Group Research*, 25(1):105–136, 1994.

[30] Miranda, S. M. and Saunders, C. Group Support Systems: An Organization Development Intervention to Combat Groupthink. *Public Administration Quarterly*, 19(2):193–216, 1995.

[31] Moorhead, G., Ference, R., and Neck, C. P. Group Decision Fiascoes Continue: Space Shuttle Challenger and a Revised Groupthink Framework. *Human Relations*, 44(6):539–550, June 1991.

[32] Munneke, L., Andriessen, J., Kanselaar, G., and Kirschner, P. Supporting interactive argumentation: Influence of representational tools on discussing a wicked problem. *Computers in Human Behavior*, 23(3):1072–1088, May 2007.

[33] Neundorf, K. *The Content Analysis Guidebook*. SAGE Publications Inc., Thousand Oaks, CA, 2002.

[34] Nunamaker, J. F., Dennis, A. R., Valacich, J. S., Vogel, D., and George, J. F. Electronic meeting systems. *Communications of the ACM*, 34(7):40–61, July 1991.

[35] Paasche-Orlow, M. K., Taylor, H. A., and Brancati, F. L. Readability standards for informed-consent forms as compared with actual readability. *The New England Journal of Medicine*, 348(8):721–726, 2003.

[36] Park, W.-w. A Review of research on Groupthink. *Journal of Behavioral Decision Making*, 3(4):229–245, Oct. 1990.

[37] Redish, J. Readability formulas have even more limitations than Klare discusses. *ACM Journal of Computer Documentation*, 24(3):132–137, 2000.

[38] Ritchey, T. *Wicked Problems - Social Messes: Decision Support Modelling With Morphological Analysis*. Risk, Governance and Society. Springer, 2011.

[39] Rogers, R., Harrison, K. S., Shuman, D. W., Sewell, K. W., and Hazelwood, L. L. An analysis of Miranda warnings and waivers: comprehension and coverage. *Law and human behavior*, 31(2):177–92, Apr. 2007.

[40] Rose, J. D. Diverse Perspectives on the Groupthink Theory â A Literary Review. *Emerging Leadership Journeys*, 4(1):37–57, 2011.

[41] Schwarz, B. B., Neuman, Y., Gil, J., and Ilya, M. Construction of Collective and Individual Knowledge in Argumentative Activity. *Journal of the Learning Sciences*, 12(2):219–256, Apr. 2003.

[42] Smith, S. Groupthink and the hostage rescue mission. *British Journal of Political Science*, 15:453–458, 1984.

[43] van Gelder, T. Argument mapping with reason! able. *The American Philosophical Association Newsletter on Philosophy and Computers*, pages 85–90, 2002.

[44] van Gelder, T. Enhancing Deliberation Through Computer Supported Argument Mapping. In Kirschner, P. A., Shum, S. J. B., and Carr, C. S., editors, *Visualizing Argumentation: Software Tools for Collaborative and Educational Sense-Making*, chapter 6, pages 97–115. Springer, London, 2003.

[45] Wagner, C., Cheung, K. S., Ip, R. K., and Bottcher, S. Building Semantic Webs for e-government with Wiki technology. *Electronic Government, an International Journal*, 3(1):36–55, 2006.

[46] West, J. A. and West, M. L. *Using wikis for online collaboration: The power of the readâwrite web*. John Wiley & Sons., San Francisco, CA, 2009.

[47] Wimmer, R. D. and Dominick, J. R. *Mass media research*. Wadsworth, Belmont, CA, 1997.