Automating Detection of Framing Agency in Design Team Talk

Dr. Ardeshir Raihanian Mashhadi, University at Buffalo, SUNY

Dr. Ardeshir Raihanian is an assistant professor of teaching in the Department of Mechanical and Aerospace Engineering at University at Buffalo. His research interests include user-centric design, sustainable design, user behavior simulation and agent based modeling. He also researches and publishes in areas surrounding engineering education. He has won multiple awards, including Design for Manufacture and the Life Cycle Technical Committee Best Paper(2017) and the International Life Cycle Academy Award for the best paper on Sustainable Consumption (2017). He is also responsible for teaching introductory, intermediate and advanced design related courses in the Department of Mechanical and Aerospace Engineering at University at Buffalo.

Dr. Vanessa Svihla, University of New Mexico

Dr. Vanessa Svihla is a learning scientist and associate professor at the University of New Mexico in the Organization, Information and Learning Sciences program and in the Chemical and Biological Engineering Department. She served as Co-PI on an NSF RET Grant and a USDA NIFA grant, and is currently co-PI on three NSF-funded projects in engineering and computer science education, including a Revolutionizing Engineering Departments project. She was selected as a National Academy of Education / Spencer Postdoctoral Fellow and a 2018 NSF CAREER awardee in engineering education research. Dr. Svihla studies learning in authentic, real world conditions; this includes a two-strand research program focused on (1) authentic assessment, often aided by interactive technology, and (2) design learning, in which she studies engineers designing devices, scientists designing investigations, teachers designing learning experiences and students designing to learn.

Automating Detection of Framing Agency in Design Team Talk

Abstract

Those who teach design contend with issues of authenticity and engagement. A problem that is too narrow or open can be challenging for students, yet finding that Goldilocks middle ground is complicated by many factors. Framing agency—making decisions that are consequential to framing and solving design problems—appears to provide clarity about student engagement with different types of design problems. However, detecting framing agency in design team talk is a labor intensive process. The purpose of this study was to evaluate a variety of text mining approaches for suitability in detecting framing agency in transcribed talk. We found no correlation between human-coding and sentiment analysis. However, interestingly, polarity derived from sentiment analysis did differentiate between a team that displayed almost no framing agency and those that did, with the former showing a high level of positivity. This reflects the lack of struggle, high certainty and agreeability the team displayed as they quickly agreed they all had similar and therefore correct answers. We also trained a Regularized Support Vector Machine Classifier to predict levels of framing agency, with the human-coded data as training data. The model showed 89% accuracy in detecting high framing agency. Given the recent increase in quality of auto-transcription tools, such approaches may lead to in-situ detectors in future.

Introduction and research purpose

Faculty who teach design face many instructional decisions, from identifying design problems that are realistic enough, open-ended enough, and feasible, to forming and scaffolding teams to learn design practices. Chief amongst these challenges is supporting students to collaboratively frame design problems. While some teams seem to intuitively understand how to direct and negotiate this process, others struggle—productively or unproductively. Recent research characterized this process with a new construct, *framing agency*, which focuses on students making decisions that are consequential to framing the design problem and how they learn as they direct their design work.

Using discourse analysis, presence or absence of framing agency can be detected in team talk. However, this analysis is slow and labor intensive. The purpose of this study was to explore and evaluate the potential of various text analysis and data mining tools as a means to more rapidly process transcripts of design team talk. Specifically, we would like to address the following research questions:

> RQ1 - How distinctive is the usage of words among groups who show different levels of framing agency? In particular, is the distribution of the words used by groups who had shown different levels of framing agency in the discourse analysis different?
>
> RQ2 - To what extent can the sentiment structure of the group discussions be used as a proxy for framing agency?
>
> RQ3 - Can the calculation of framing agency scores be automated using a supervised learning approach?

Theoretical Framework

Broadly, we situate our work as constructionist a view of learning that fits well with design learning because of its emphasis on constructing meaning in a public facing way by addressing realistic and relevant problems [1]. A key feature of such design learning environments is that learners have agency—they make decisions. However, the scope of decisions they may make varies by setting, from kit-based formulaic design challenges with limited scope, to underdefined entrepreneurial design with broad scope. To understand the nature of their decision making—that is, the nature of their agency—we use the notion of *framing agency*, defined as opportunities to make decisions that are consequential to the problem frame and how designers go about learning and solving the problem. We build on prior research on ownership, agency, learning, and design problem framing.

Designers make consequential decisions as they frame problems

Because design problems have many possible solution paths and solutions, designers must make many decisions, not just about whether a solution will meet needs or is feasible, but about what the design problem actually is. This process has been called by various names, including problem scoping, framing, problem definition, and others. While some define differences between these terms, close review finds a high degree of overlap in the activities and purposes. For instance, some scholars reference the findings of Atman and colleagues (who prominently refer to *scoping*) using the terms framing and scoping interchangeably [2]. Likewise, Atman and colleagues reference work by those who exclusively discuss *framing* [3-5], yet refer to that work as *scoping*. Influenced by Schön's [6] view of design as a reflective conversation with materials, we use the word he commonly used—framing [7], though we are influenced by work using others terms.

In framing design problems, designers make many decisions that are consequential to the problem. They decide what to include and exclude from the problem, bounding it [8]. To do so, they gather information to fill in gaps in their understanding [9]. Experienced, skillful designers engage in framing and reframing deliberately and repeatedly, throughout design process [3, 10-13]. They pay attention to the customer/stakeholder needs, logistics, and constraints [14]. Thus, they make decisions not just about the problem, but also about how to proceed. This is consequential to both what they learn about the problem and how they learn it.

To understand this form of directed decision making, we turn to agency theory, which posits that agency comprises the opportunity to make decisions, the act of making decisions, and the outcome [15]. Placing agency into the context of design, we term this *framing agency* to differentiate making consequential decisions about how to frame a design problem from other forms of agency (e.g., deciding to check email, deciding to say no to a new project, etc.). This is particularly helpful for understanding the kinds of experiences learners may benefit from in developing as designers.

Where is agency in design education?

Supporting learners to develop capacity to frame problems is like teaching them any other skill, in that it requires understanding of why the skill matters and opportunities to practice and reflect on performance [16]. But because of the nature of design, it is also very different to manage this kind of learning. One way educators have tackled this issue is to constrain the problem such that there is a correct answer. However, such deterministic problems do little to help learners develop the capacity to direct their own framing of problems. Indeed, because students have typically encountered many such problems in prior coursework, they may struggle with the ambiguity presented by problem framing [17]. Even when facing an industry-sponsored capstone project, they may treat the problem as having a single correct answer [18]. Research suggests that when students are supported to make consequential decisions, they feel a greater sense of ownership over their work [19, 20]. Framing agency, therefore, can serve as a lens into whether students are learning to negotiate the process of framing design problems. In our past work, we found that students' talk in their design teams was indicative of whether they treated the problems as framed for them and not open to reframing, or as problems they themselves needed to frame [18, 21].

Methods

Research design

In order to meet our research aims, we first conducted discourse analysis and then explored data mining techniques. We provide an overview below. We then, in the results, provide a detailed account of each method used to answer the specific research question, along with the findings.

Participants, setting & data collection

We used a data corpus comprising 11 50-minute audio recordings collected in three cohorts (C1, C2, C3) of a sophomore-level design project focused on designing an algal biofuel plant for a rural community. Teams were audio recorded as they discussed ideas they had independently researched, but needed to make a team decision about.

Data analysis: Discourse analysis

All audio records were transcribed verbatim and a subset of three teams were scored for level of framing agency using discourse analytic methods [22, 23]. Past research has clarified that agency can be detected in natural talk and that such talk is reflective of thinking [20, 24]. Over iterative work with multiple datasets [18, 21], we identified the importance of pronouns and verb forms for distinguishing levels of agency (Table 1), based also in research on agency and discourse analysis [22, 23]. When speakers used passive verbs that typically indicate a lack of personal or team control over the situation, such as "had to," "needed to," or "were required to" in place of "did" or "have to" in place of "do," they were displaying low or no framing agency. In contrast, highest levels of agency, denoting personal or team control through verbs like "did" and "do" display a level of certainty that may not be warranted for early design work, except when reporting individual progress to team members. As they work together to frame problems, member framing agency is typified as shared and tentative. Speakers use modal verbs that indicate potential for personal or team control over the situation, such as "could," "would," or "should," in place of "did" or "do." We also differentiated between talk in which a speaker self-directed ("I will do that") versus directed others ("You do that"). We found very few examples of a speak directing others; in most instances, this involved a recounting of previously-provided information or of limitations imposed by others, not by the speaker. Based on this grounding in the data, we found that individual accounts ("I …") displayed higher agency than those that involved directing others ("you …). Regardless of their relative levels, the capacity to distinguish them from one another is salient in the initial discourse analysis.

This analysis highlighted that while the teams in cohorts 1 and 3 displayed high levels of framing agency, the team in cohort 2 displayed almost no framing agency; instead they used their agency to treat the problem as having a single correct answer [18, 21].

Table 1. Framing agency levels, characterized from discourse analysis

Description	***Score***	***Level***	***Locus of control***	***Certainty***
I did that / I do that	10	High	Self	Certain
I could have done that / I could do that/ let me do that	9	High	Self	Potential
You do that / You did that / implicit you "do that"	8	High	Self	Certain
You could do that / You could have done that	7	High	Self	Potential
You have to do that	6	High	Self	Certain
We did that / We do that / implicit "we"	5	Moderate	Shared	Certain
We could have done that / We could do that /Let us do that	4	Moderate	Shared	Potential
It could be / It could have been (typically a reference to problem ideas)	3	Moderate	Shared with objects	Potential
It must be / It must not be / It is (typically a reference to endemic problem requirements/constraints)	2	Low	Shared with problem	Certain
I have to do that / We have to do that	1	None	Situation or environment	Certain

Data analysis: Text mining overview

We applied various text mining techniques to analyze the transcripts. First, we conducted exploratory analysis of words in order to investigate RQ1. We then performed a sentiment analysis to investigate any possible relationship between the sentiment scores and framing agency in order to answer RQ2. Lastly, we created a prediction model using supervised learning to predict high framing agency instances using previously human-coded data (RQ3).

General text analytics framework

We labeled data from the three teams as C1, C2 and C3 and in the remainder of this paper we will refer to them accordingly. C1, C2 and C3 included a total of 2211 sentences comprising of 16125 words. The first step of the text analysis is the exploratory analysis of the frequency of words across the three subsets of data. We removed words that were highly frequent but did not add any intrinsic value to the analysis. The removed words are as follows: "oh", "inaudible", "inaud", "like", "the", "from", "but", "to", "that", "of", "it", "yes", "yeah" and "does". Some of these words like "inaudible" and "inaud" are words that were introduced in the data set because of verbatim transcription of the recordings and the rest are articles, conjunctions, etc.

For clarity, we present detailed methods of analysis paired with their results.

Results

RQ1: Exploratory analysis of word distribution

The 30 most frequent words in each subset and the corresponding word occurrences are presented in Table 2. It should be noted that C1 included a smaller data set (fewer sentences) and therefore, the overall frequency of the words is smaller in it, while C2 and C3 are more similar in terms of the total count of words. We can observe in Table 2 that the overall distribution of words across the three subsets are quite similar. The most frequent words are those that are generally more frequent in the English language (e.g., pronouns), as well as specific words referring to the design problem that was assigned to the students (e.g., reactor, lipid etc.). In order to account for the inherent frequency of words, we have also used a different measure from simple occurrences, called tf-idf that refers to term frequency-inverse document frequency [25]. Equation 1 illustrates how this measure is calculated.

The differences reflected in Table 2 suggested that sentiment analysis could provide insight. For instance, words like "okay" are prominent in both C2 and C3, but not in C1.

Table 2. Most frequent words across the three subsets

C1		**C2**		**C3**	
word	***Count***	***word***	***Count***	***word***	***Count***
a	72	so	154	so	158
i	69	we	138	i	153
you	54	and	106	we	124
so	52	i	106	you	116
and	50	a	99	a	107
it's	47	you	80	and	103
we	47	have	67	one	82
be	46	is	66	what	81
in	40	one	57	have	78
is	36	do	55	is	72
have	34	for	52	it's	69
with	30	in	52	in	61
bio	28	was	52	just	52
just	28	what	50	on	51
that's	28	ok	49	about	48
reactor	27	that's	47	okay	47
not	26	lipid	42	do	45
open	26	just	40	was	44
think	26	per	39	for	41

we're	26	our	38	if	41
for	25	this	38	can	40
are	24	rate	37	know	39
more	24	it's	36	right	39
if	23	right	34	um	39
pond	22	growth	33	are	37
an	21	on	30	that's	37
closed	21	how	29	don't	35
this	21	with	29	gonna	35
can	20	all	28	we're	34
do	20	okay	28	as	33

Assume C is a corpus consisting of a collection of documents $d_1$to d_n. Tf-idf of the word "word" $tf - idf(\text{"word"}, d_1, C)$ is calculated as follows:

$$\frac{frequency\ of\ \text{"word" in } d_1}{frequency\ of\ documents\ with\ \text{"word" in C}} = \frac{\#\ of\ \text{"word" in } d_1}{total\ \#\ of\ \text{words in } d_1} \times \frac{1}{\frac{\#\ of\ \ documents\ including\ \text{"word"}}{n}} \qquad (1)$$

Therefore, tf-idf is a measure of the level of information a word represents in a document (i.e., any collection of words). In other words, if a word is repeated several times across C (i.e., large document frequency), it is probably not very important (e.g., the word "so"). On the other hand, if a word is rarely present across C (i.e., small document frequency) but is repeated several times in a document (i.e., large term frequency), it should contain valuable information regarding that document.

Table 3 presents the first 10 words with highest values of tf-idf in the three subsets. Comparison of such words does not suggest any comparative information regarding the differences in the level of framing agency that each group of students showed. However, these words do provide some insight into how each team approached framing the problem. Cohort 1 began with a simpler task, choosing between an open pond or bioreactor for growing algae. Despite this narrower decision, they considered many factors, including which provided more room for innovation and which might make lipid extraction simpler or more difficult. The teams in cohorts 2 and 3 were asked to first make their choices regarding which kinds of algae to grow, and they were encouraged to individually explore many options and factors prior to choosing. In the case of cohort 2, they considered few strains and few factors, and quickly affirmed their agreement with one another, and this is reflected in Table 3. In contrast, the team in cohort 3 discussed a wide range of algae (including some less common strains, referred to in shorthand by students as broni and groni) and factors. Thus, these results provide insight into the breadth of information students used to inform their problem framing work.

Table 3. Tf-idf scores of words across the three subsets.

C1				
word	n	tf	idf	tf_idf
innovation	1	1	4.779123	4.779123
extraction	1	1	4.373658	4.373658
exactly	1	1	4.085976	4.085976
okay	1	1	3.392829	3.392829
too	1	1	3.169686	3.169686
looks	1	0.5	5.472271	2.736135
sense	1	0.5	5.472271	2.736135
gosh	1	0.5	5.472271	2.736135
tubes	1	0.5	5.472271	2.736135
moving	1	0.5	5.472271	2.736135
C2				
boarding	1	1	6.270988	6.270988
concentration	1	1	6.270988	6.270988
either	1	1	6.270988	6.270988
hmmmm	1	1	6.270988	6.270988
nanochloroxine	1	1	6.270988	6.270988
god	1	1	6.270988	6.270988
duh	1	1	6.270988	6.270988
riight	1	1	6.270988	6.270988
ya	1	1	6.270988	6.270988
apparently	1	1	6.270988	6.270988
C3				
ethynol	1	1	6.274762	6.274762
barely	1	1	6.274762	6.274762
chraneortrac	1	1	6.274762	6.274762
criterias	1	1	6.274762	6.274762
goni	1	1	6.274762	6.274762
gronia	1	1	6.274762	6.274762
hallelujah	1	1	6.274762	6.274762
talyor	1	1	5.581615	5.581615
sara	1	1	5.581615	5.581615
broni	1	1	5.581615	5.581615

The analysis of bi-grams reveals more fruitful insights. We have looked at bi-grams (i.e., 2-word combinations that occur together) across the three subsets and sorted them based on the number of times they appeared. Table 4 summarizes the results. Our discourse analysis showed that C1 and C3 were groups who generally showed significantly higher levels of framing agency. This pattern can be observed in the frequency of bi-grams as well. As an example, in both C1 and C3 the phrase

“I think”, which associates with a high level of framing agency, is the most frequent bi-gram. In contrast, this bigram was less frequent in the C2, the group that did not display framing agency. One of their most common bigrams was “yeah yeah” which also reflects their agreeability, again, suggesting that sentiment analysis may be of use.

Table 4. Bi-gram frequency across the three subsets.

C1		**C2**		**C3**	
bigram	***Frequency***	***bigram***	***Frequency***	***bigram***	***Frequency***
i think	24	growth rate	27	i think	27
bio reactor	23	yeah yeah	24	i don't	26
a lot	20	do you	21	what about	23
have to	19	per day	20	have to	22
you can	19	ok so	17	of the	20
in the	15	per liter	17	don't know	18
lot of	15	for the	15	growth rate	18
of the	15	like a	15	we have	18
open pond	15	lipid content	15	like a	17
to be	14	so we	15	okay so	16
closed system	13	we have	15	yeah yeah	16
the bio	13	and then	14	in the	15
would be	13	heh heh	14	that one	15
you have	13	do we	12	if you	14
in a	12	i think	12	on the	14
an open	11	a 2	11	so we	14
it's not	11	huh huh	11	you know	14
the algae	11	i don't	11	do you	13
and then	10	we should	11	are we	12
we have	10	you guys	11	it was	12
all the	9	going to	10	you have	12
bio reactors	9	kind of	10	NA	12
inaudible 00	9	milligrams per	10	cost efficiency	11
one of	9	most important	10	i mean	11
so i	9	that we	10	it has	11
the open	9	to be	10	like the	11
00 11	8	want to	10	you can	11
and you	8	we need	10	a one	10
it would	8	we want	10	did you	10
like the	8	a high	9	do we	10

The exploratory analysis of the bi-grams suggests that the use of word patterns may be useful in capturing the framing agency.

RQ2. Sentiment analysis

Given findings in exploratory analysis, we wondered if there was a relationship between the sentiment structure of the sentences and the framing agency level. In other words, can the sentiment score of a sentence act as a proxy for framing agency? In general, sentiment analysis utilizes a dictionary that assigns a sentiment score to certain words and then calculates a sentiment score for a sentence or a phrase based on the scores of the included words. We have used two lexicons for sentiment analysis. The first one is based on the *Bing* dictionary [26]. This method assigns a polarity score between -1 (negative) and +1 (positive) and also recognizes negators (words that reverse the polarity of a word) and amplifiers (words that change the intensity of polarity of words). The second dictionary we used is the NRC lexicon [27]. In this method, the presence of eight different emotions and their associated valence is calculated. Our goal was to investigate the extent to which the sentiment structure of the group discussions can be used as a proxy for framing agency. Therefore, we studied the Pearson correlation coefficients of the human-coded framing agency levels and the sentiment scores. The human-coded levels were on a scale of 1-10, where 10 refers to the highest level of framing agency. Table 5 summarizes the bivariate Pearson correlation analysis results of C1 as an example. 'Code' refers to the human-coded framing agency level, 'polarity' is the sentiment score based on the Bing lexicon and the rest of the columns refer to the eight different emotions captured using the NRC lexicon.

Table 5. Correlation analysis between framing agency and sentiment scores.

code	***anger***	***anticipation***	***disgust***	***fear***	***joy***	***sadness***	***surprise***	***trust***	***negative***	***positive***	***polarity***
1.00	-0.04	-0.11	-0.03	-0.05	-0.06	-0.03	-0.02	-0.08	-0.11	-0.10	-0.08
-0.04	1.00	0.15	0.34	0.46	0.14	0.36	0.26	0.19	0.37	0.15	-0.07
-0.11	0.15	1.00	-0.03	0.00	0.61	0.02	0.45	0.50	0.09	0.60	0.23
-0.03	0.34	-0.03	1.00	0.42	-0.05	0.35	0.14	0.12	0.32	0.01	-0.06
-0.05	0.46	0.00	0.42	1.00	-0.02	0.60	0.08	0.07	0.53	0.03	-0.12
-0.06	0.14	0.61	-0.05	-0.02	1.00	0.00	0.53	0.58	-0.01	0.62	0.18
-0.03	0.36	0.02	0.35	0.60	0.00	1.00	0.04	0.03	0.64	-0.03	-0.15
-0.02	0.26	0.45	0.14	0.08	0.53	0.04	1.00	0.49	0.10	0.45	0.16
-0.08	0.19	0.50	0.12	0.07	0.58	0.03	0.49	1.00	0.06	0.62	0.27
-0.11	0.37	0.09	0.32	0.53	-0.01	0.64	0.10	0.06	1.00	0.10	-0.14
-0.10	0.15	0.60	0.01	0.03	0.62	-0.03	0.45	0.62	0.10	1.00	0.28
-0.08	-0.07	0.23	-0.06	-0.12	0.18	-0.15	0.16	0.27	-0.14	0.28	1.00

As can be seen in Table 5, we found no correlation between human-coding and sentiment analysis. Although we saw differences in words that suggested sentiment differences by cohort, this did not reflect framing agency. This is sensible, as talk that sentiment analysis would code as negative (e.g., "I don't think we want to use that extraction method") versus as positive (e.g., "I do think we want to use that extraction method") display the same level of framing agency. However, interestingly, polarity derived from sentiment analysis did differentiate between a team that displayed almost no framing agency and those that did, with the former showing a high level of positivity. This reflects the lack of struggle, high certainty and agreeability the team displayed as they quickly agreed they all had similar and therefore correct answers. Figure 1 illustrates the polarity of the conversations chronologically across the three teams. The x-axis represents the line ID (each line is a team member talking in their turn) and the y-axis represents the polarity score calculated based on the number of positive words subtracted by the number of negative words at each turn using the Bing dictionary. Interestingly, the second group C2, for which the discourse analysis showed the lowest level of framing agency, has a higher polarity score on average.

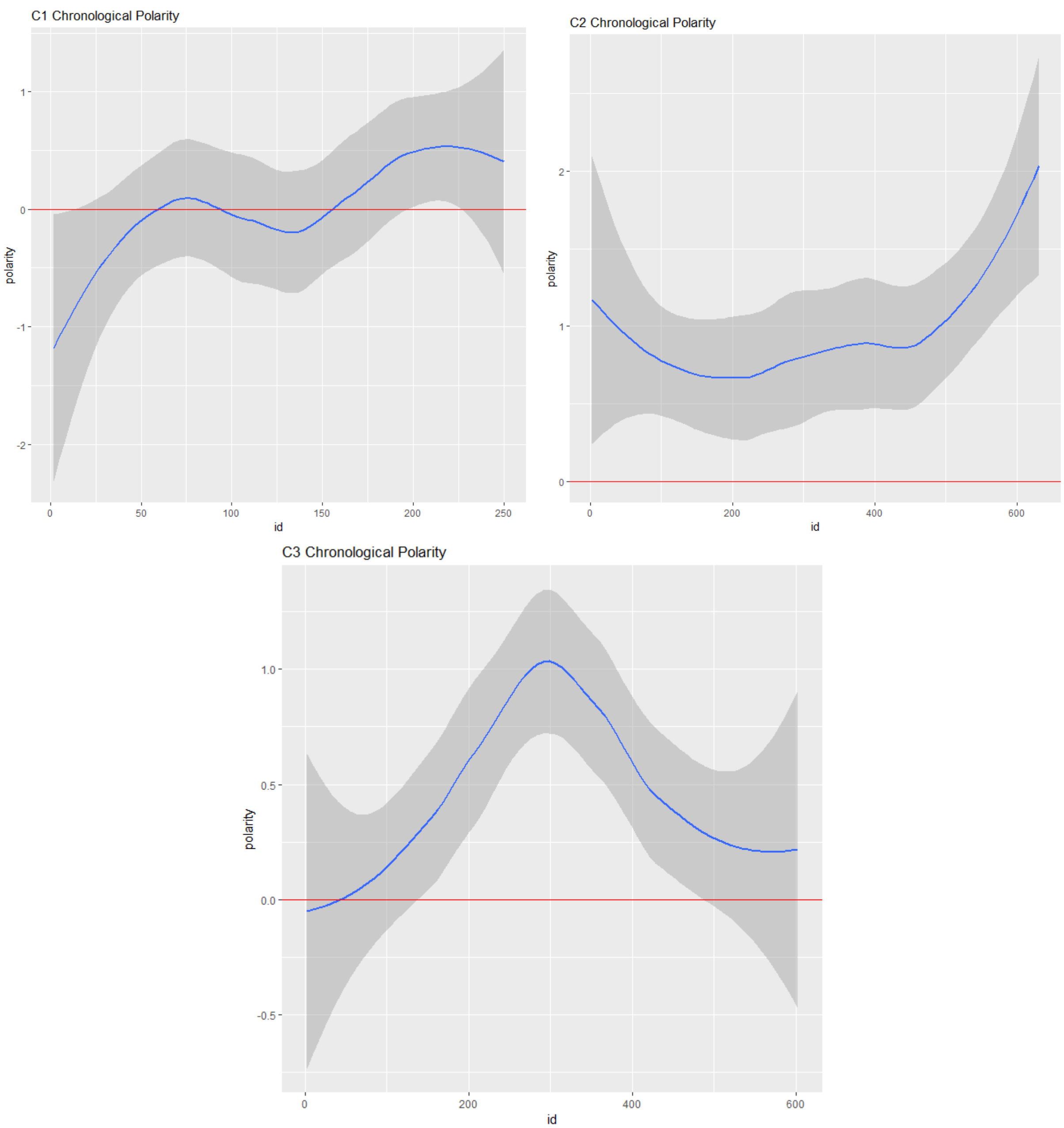

Figure 1. Chronological polarity analysis for C1 (Top, left), C2 (Top, right), C3 (Bottom)

RQ3. Predicting framing agency using text classification

The third approach we took was to investigate the possibility of automating the calculation of framing agency scores using a supervised learning approach. If successful, this could help in identification of low (or high) levels of framing agency without extensive and tedious human coding.

Here we used a “bag of words” approach for text analysis. The bag of words representation, which has been heavily used in information retrieval research, assumes that words and their frequency of occurrence in a text, without paying attention to the ordering of them, could be a good representation of the text for many tasks [28, 29]. For the sake of having a larger set of data to train the supervised learning model, we broke the conversations into sentences and combined the textual information we had from the three teams into one corpus. Note that here we do not aim to compare anything across the teams. The goal was to build a supervised learning model that, if given a new sentence, can predict its level of framing agency with reasonable accuracy.

The first step of the analysis is to preprocess the textual information through a data cleaning framework. Therefore, the textual data was transformed to lowercase, the abbreviations and contractions were replaced by the complete terms and numbers and punctuations were removed from the texts. Then the sentences were tokenized (converted to a bag of words). Also, a dictionary of redundant words that were introduced during the transcription process ("inaudible","inaud") was created and the corresponding words were removed from the data. Note that the common stop words were kept in the data as the discourse analysis suggested that they may have a value in identifying the level of framing agency. Table 6 illustrates an example of the preprocessed text and the corresponding conversion using the bag of words representation.

Table 6. Example of tokenizing using a bag of words approach. Words with less than 3 letters are not shown.

Words / Text	are	costs	harve	high	one	open	pond	probl	says	see	that	the	too	using	you
you see that is one of the problems of using an open pond	0	0	0	0	1	1	1	1	0	1	1	1	0	1	1
it says that the harvesting costs are too high	1	1	1	1	0	0	0	0	1	0	1	1	1	0	0

We labeled the framing agency scores as follows:

Class Label	Human-coded Framing Agency Level	Description
0	0	Uncodable
1	1 ≤ < 8	Baseline
2	≤ 8	High level of framing agency

We treated each sentence as an observation. Therefore, the number of occurrence of each unique word would act as predictors and the response variable was the class label. Therefore, a classification model could be trained. In order to avoid overfitting, the following steps were taken: First, the dataset was randomly split into train and validation sets (2/3 train, 1/3 validation); second, the model was trained using 10-fold cross validation with 5 repeats; third, the prediction accuracy of the model was tested using the validation set.

We trained a Support Vector Machine (SVM) classifier for text classification. SVM is a supervised learning algorithm that learns a decision boundary to classify data by maximizing the margin between the training patterns [30]. SVM has been shown to be very well suited for text categorization [29]. Tables 7 and 8 summarize the prediction results of the trained model. Table 7 presents the confusion matrix of the prediction model depicting the prediction results against the actual labels of the validation data in each class. Table 8 provides classification-specific measures of accuracy: (1) sensitivity provides a measure of proportion of actual positive cases that were predicted; (2) specificity provides a measure of actual negative cases that were predicted; and (3) balanced accuracy is the average of sensitivity and specificity. In other words:

$$Sensitivity = \frac{True\ Positive}{True\ Positive + False\ Negative} \tag{2}$$

$$Specificity = \frac{True\ Negative}{True\ Negative + False\ Positive} \tag{3}$$

No information rate represents the largest proportion of the observed classes. The reported p-value is according to a hypothesis test that checks if the overall accuracy of the prediction is higher than the largest proportion of classes.

As we can see, the trained model predicted the right class with almost 90% accuracy, which is a very good accuracy level in text mining [31]. In other words, the trained model can accurately identify if a given sentence demonstrates the highest level of framing agency.

Table 7. Confusion matrix of the classification model.

Prediction	Reference		
	Class: 0	Class: 1	Class: 2
Class: 0	142	28	5
Class: 1	13	364	19
Class: 2	0	16	165

Table 8. Statistics by class for the classification model.

	Class: 0	Class: 1	Class: 2
Sensitivity	0.9161	0.8922	0.8730
Specificity	0.9447	0.9070	0.9716
Balanced Accuracy	0.9304	0.8996	0.9223
Overall Statistics: Accuracy : 0.8923, No Information Rate : 0.5426 P-Value [Acc > NIR] : < 2e-16			

Significance and implications

We explored three primary techniques to automate the laborious analysis of conversational data, specifically aiming to detect levels of framing agency in design team talk. While our first two approaches did not identify framing agency, they did shed light on ongoing areas of study in design team talk.

First, exploratory analysis, in particular tf-idf scores of words across three cohorts provided insight into how expansively teams were framing design problems. Given that past research contrasting experienced and novice designers has suggested that novices tend to frame problems too narrowly [3, 10-13], future work could investigate this technique as a means to diagnose this narrowness and explore whether it relates to design outcomes.

Second, while sentiment analysis also did not serve as a proxy for framing agency, it did serve as a means to identify a malfunctioning team (C2), in which members spent time agreeing with one another about their narrowly defined problem. In fact, that it did not serve as a proxy for framing agency validates our earlier characterization of framing agency, as both positive and negative sentiments can be found at the same level of agency. Future work can build on this to investigate whether sentiment analysis may be a means to detect groupthink in design teams [32, 33].

Third, a supervised learning approach did serve as a proxy for framing agency with a high degree of accuracy. However, our approach at this point, given the scope of our dataset, was to restrict the number of levels in framing agency.

Our ongoing work explores the extensibility of this approach to new datasets, while also tuning our methods to improve accuracy, including detecting more levels of framing agency. Given the recent increase in quality of auto-transcription tools, such approaches may lead to in-situ detectors using cloud-based Natural Language Processing platforms in future. Such tools could allow faculty to respond to teams struggling to make design decisions. Likewise, such technology could lead to the development of tools that help students become aware of how their talk reflects and shapes their thinking about their design work. Finally, by making it simpler to detect framing agency, or its lack, in design team conversations, faculty who teach design may be able to learn more about the kinds of design problems (e.g., client driven, kit-based, small scale, entrepreneurial, etc.), features of design practices that may be enhanced or omitted (e.g., time/effort spent on customer discovery or needs assessment, agile or iterative methods, etc.), and the kinds of supports new designers might most benefit from.

While significant work is yet to be undertaken to recognize such goals, and research is needed to better understand a breadth of design situations and successes, as long as this is limited to labor intensive human analysis, the breadth will remain to limited for more generalizable tools.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. EEC 1751369. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

[1] S. Papert and I. Harel, "Situating constructionism," in *Constructionism*New York, NY: Ablex Publishing Corporation, 1991, pp. 1-11.

[2] C. Williams, J. Gero, Y. Lee, and M. Paretti, "Exploring the Effect of Design Education on the Design Cognition of Sophomore Engineering Students," in *Proceedings of ASME 2011 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*Washington, DC: ASME, 2011, pp. 607-614.

[3] D. A. Schön, "Designing: Rules, types and words," *Design studies,* vol. 9, no. 3, pp. 181-190, 1988.

[4] N. Cross and A. Clayburn Cross, "Expertise in engineering design," *Research in Engineering Design,* vol. 10, no. 3, pp. 141-149, 1998.

[5] N. Cross, "Design cognition: Results from protocol and other empirical studies of design activity," in *Design knowing and learning: Cognition in design education*, vol. 5, C. M. Eastman, W. M. McCracken, and W. C. Newstetter, Eds. Oxford, UK: Elsevier Science, 2001, pp. 79-103.

[6] D. A. Schön, *The Reflective Practitioner: How Professionals Think in Action*. New York: Basic Books, 1983.

[7] D. A. Schön, "Problems, frames and perspectives on designing," *Design Studies,* vol. 5, no. 3, pp. 132-136, 1984.

[8] C. J. Atman, K. Yasuhara, R. S. Adams, T. J. Barker, J. Turns, and E. Rhone, "Breadth in Problem Scoping: a Comparison of Freshman and Senior Engineering Students," *International Journal of Engineering Education,* vol. 24, no. 2, pp. 234-245, 2008.

[9] C. J. Atman, J. Chimka, K. M. Bursic, and H. L. Nachtmann, "A comparison of freshman and senior engineering design processes," *Design Studies,* vol. 20, no. 2, pp. 131-152, 1999.

[10] M. Zahedi, L. Heaton, M. Guité, G. De Paoli, and M. Reumont, "Exploring framing within a team of industrial design students," 2016: DRS2016, Design Research Society 50th Anniversary Conference, Future-Focused Thinking, Brighton, UK.

[11] R. Valkenburg and K. Dorst, "The reflective practice of design teams," *Design Studies,* vol. 19, no. 3, pp. 249-271, 1998.

[12] S. Ahmed, K. M. Wallace, and L. T. Blessing, "Understanding the differences between how novice and experienced designers approach design tasks," *Research in Engineering Design,* vol. 14, no. 1, pp. 1-11, 2003.

[13] M. Mehalik and C. Schunn, "What constitutes good design? A review of empirical studies of design processes," *International Journal of Engineering Education,* vol. 22, no. 3, p. 519, 2007.

[14] L. L. Bogusch, J. Turns, and C. J. Atman, "Engineering design factors: how broadly do students define problems?," in *Frontiers in Education Conference*, Kansas City, MO, 2000, vol. 2.

[15] D. Narayan and P. Petesch, "Agency, opportunity structure, and poverty escapes," *Moving Out of Poverty: Cross-Disciplinary Perspectives on Mobility,* pp. 1-44, 2007.

[16] D. A. Kolb, *Experiential learning: Experience as the source of learning and development*. FT press, 2014.

[17] M. E. Jordan and R. R. McDaniel Jr, "Managing uncertainty during collaborative problem solving in elementary school teams: The role of peer influence in robotics engineering activity," *Journal of the Learning Sciences,* vol. 23, no. 4, pp. 490-536, 2014.

[18] V. Svihla, T. B. Peele-Eady, and Y. Chen, "Agency in framing design problems " in *AERA Annual Meeting*, Toronto, ON, Canada, 2019.

[19] D. I. Hanauer, J. Frederick, B. Fotinakes, and S. A. Strobel, "Linguistic analysis of project ownership for undergraduate research experiences," *CBE-Life Sciences Education,* vol. 11, no. 4, pp. 378-85, Winter 2012.

[20] A. Haapasaari, Y. Engeström, and H. Kerosuo, "The emergence of learners' transformative agency in a Change Laboratory intervention," *Journal of Education and Work,* vol. 29, no. 2, pp. 232-262, 2016.

[21] V. Svihla, J. R. Gomez, M. A. Watkins, and T. B. Peele-Eady, "Characterizing framing agency in design team discourse," in *Proceedings of the ASEE 126th Annual Conference and Exhibition*: ASEE, 2019.

[22] J. P. Gee, *An introduction to discourse analysis: Theory and method*. Routledge, 2014.

[23] A. W. Konopasky and K. M. Sheridan, "Towards a Diagnostic Toolkit for the Language of Agency," *Mind, Culture, and Activity,* vol. 23, no. 2, pp. 108-123, 2016.

[24] L. M. Ahearn, "Language and agency," *Annual review of anthropology,* pp. 109-137, 2001.

[25] J. Ramos, "Using tf-idf to determine word relevance in document queries," in *Proceedings of the first instructional conference on machine learning*, 2003, vol. 242, pp. 133-142: Piscataway, NJ.

[26] M. Hu and B. Liu, "Mining opinion features in customer reviews," in *AAAI*, 2004, vol. 4, no. 4, pp. 755-760.

[27] S. M. Mohammad and P. D. Turney, "Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon," in *Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text*, 2010, pp. 26-34: Association for Computational Linguistics.

[28] T. Joachims, "A Probabilistic Analysis of the Rocchio Algorithm with TFIDF for Text Categorization," in *International Conference on Machine Learning*, 1997.

[29] T. Joachims, "Text categorization with support vector machines: Learning with many relevant features," in *European conference on machine learning*, 1998, pp. 137-142: Springer.

[30] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the fifth annual workshop on Computational learning theory*, 1992, pp. 144-152.

[31] N. Archak, A. Ghose, and P. G. Ipeirotis, "Deriving the pricing power of product features by mining consumer reviews," *Management science,* vol. 57, no. 8, pp. 1485-1509, 2011.

[32] I. L. Janis, *Groupthink*. Houghton Mifflin Boston, 1982.

[33] W. W. Park, "A review of research on groupthink," *Journal of Behavioral Decision Making,* vol. 3, no. 4, pp. 229-245, 1990.