Supplementary Material for What Gets Echoed? Understanding the "Pointers" in Explanations of Persuasive Arguments

David Atkinson and **Kumar Bhargav Srinivasan** and **Chenhao Tan**

Department of Computer Science
University of Colorado Boulder
Boulder, CO
`david.i.atkinson, kumar.srinivasan, chenhao.tan@colorado.edu`

A Supplemental Material

A.1 Preprocessing.

Before tokenizing, we pass each OP, PC, and explanation through a preprocessing pipeline, with the following steps:

1. Occasionally, `/r/ChangeMyView`'s moderators will edit comments, prefixing their edits with "Hello, users of CMV" or "This is a footnote" (see Table 1). We remove this, and any text that follows on the same line.

2. We replace URLs with a "@url@" token, defining a URL to be any string which matches the following regular expression: `(https?://[^\s)]*)`.

3. We replace "Δ" symbols and their analogues—such as "δ", "&;#8710;", and "!delta"—with the word "delta". We also remove the word "delta" from explanations, if the explanation starts with delta.

4. Reddit–specific prefixes, such as "u/" (denoting a user) and "r/" (denoting a subreddit) are removed, as we observed that they often interfered with spaCy's ability to correctly parse its inputs.

5. We remove any text matching the regular expression `EDIT(.*?):.*` from the beginning of the match to the end of that line, as well as variations, such as `Edit(.*?):.*`.

6. Reddit allows users to insert blockquoted text. We extract any blockquotes and surround them with standard quotation marks.

7. We replace all contiguous whitespace with a single space. We also do this with tab characters and carriage returns, and with two or more hyphens, asterisks, or underscores.

Sample footnote: Hello, users of CMV! This is a footnote from your moderators. Wed just like to remind you of a couple of things. Firstly, please remember to read through our rules. If you see a comment that has broken one, it is more effective to report it than downvote it. Speaking of which, *downvotes dont change views**! If you are thinking about submitting a CMV yourself, please have a look through our **popular topics wiki first. Any questions or concerns? Feel free to message us**. Happy CMVing!*
Sample subreddit reference: r/ideasforcmv, /r/nba
Sample URL : https://www.quora.com/profile/
Sample user reference: u/Ansuz07
Sample edit: EDIT for clarification: This isn't to suggest that you have to remain financially independent to vote

Table 1: Sample data that were affected by preprocessing.

Tokenizing the data. After passing text through our preprocessing pipeline, we use the default spaCy pipeline to extract part-of-speech tags, dependency tags, and entity details for each token[1] (Honnibal and Montani, 2017). In addition, we use NLTK to stem words (Loper and Bird, 2002). This is used to compute all word level features discussed in Section 4 of the main paper.

A.2 PC Echoing OP

Figure 1b shows a similar U-shape in the probability of a word being echoed in PC. However, visually, we can see that rare words seem more likely to have high echoing probability in explanations, while that probability is higher for words with moderate frequency in PCs. As PCs tend to be longer than explanations, we also used the echoing probability of the most frequent words to normalize the probability of other words so that they are comparable. We indeed observed a higher likelihood of echoing the rare words, but lower likelihood of echoing words with moderate frequency

---

[1] We ignore all tokens tagged as "SPACE" by the part of speech tagger.

in explanations than in PCs.

A.3 Feature Calculation

Given an OP, PC, and explanation, we calculate a 66–dimensional vector for each unique stem in the concatenated OP and PC. Here, we describe the process of calculating each feature.

1. Inverse document frequency: for a stem s, the inverse document frequency is given by $\log \frac{N}{\mathrm{df}_s}$, where N is the total number of documents (here, OPs and PCs) in the training set, and df_s is the number of documents in the training data whose set of stemmed words contains s.

2. Stem length: the number of characters in the stem.

3. Wordnet depth (min): starting with the stem, this is the length of the minimum hypernym path to the synset root.

4. Wordnet depth (max): similarly, this is the length of the maximum hypernym path.

5. Stem transfer probability: the percentage of times in which a stem seen in the explanandum is also seen in the explanation. If, during validation or testing, a stem is encountered for the first time, we set this to be the mean probability of transfer over all stems seen in the training data.

6-21. OP part–of–speech tags: a stem can represent multiple parts of speech. For example, both "traditions" and "traditional" will be stemmed to "tradit." We count the percentage of times the given stem appears as each part–of–speech tag, following the Universal Dependencies scheme (Schuster and Manning, 2016).[2] If the stem does not appear in the OP, each part–of–speech feature will be $\frac{1}{16}$.

22-24. OP subject, object, and other: Given a stem s, we calculate the percentage of times that s's surface forms in the OP are classified as subjects, objects, or something else by SpaCy. We follow the CLEAR guidelines, (Center for Language and Information Research, 2016) and use the following tags to indicate

---
[2]Note that, for English, spaCy does not use the `SCONJ` tag.

a subject: `nsubj`, `nsubjpass`, `csubj`, `csubjpass`, `agent`, and `expl`. Objects are identified using these tags: `dobj`, `dative`, `attr`, `oprd`. If s does not appear at all in the OP, we let subject, object, and other each equal $\frac{1}{3}$.

25. OP term frequency: the number of times any surface form of a stem appears in the list of tokens that make up the OP.

26. OP normalized term frequency: the percentage of the OP's tokens which are a surface form of the given stem.

27. OP # of surface forms: the number of different surface forms for the given stem.

28. OP location: the average location of each surface form of the given stem which appears in the OP, where the location of a surface form is defined as the percentage of tokens which appear after that surface form. If the stem does not appear at all in the OP, this value is $\frac{1}{2}$.

29. OP is in quotes: the number of times the stem appears in the OP surrounded by quotation marks.

30. OP is entity: the percentage of tokens in the OP that are both a surface form for the given stem, and are tagged by SpaCy as one of the following entities: `PERSON`, `NORP`, `FAC`, `ORG`, `GPE`, `LOC`, `PRODUCT`, `EVENT`, `WORK_OF_ART`, `LAW`, and `LANGUAGE`.

31-55. PC equivalents of features 6-30.

56. In both OP and PC: 1, if one of the stem's surface forms appears in both the OP and PC. 0 otherwise.

57. # of unique surface forms in OP: for the given stem, the number of surface forms that appear in the OP, but not in the PC.

58. # of unique surface forms in PC: for the given stem, the number of surface forms that appear in the PC, but not in the OP.

59. Stem part–of–speech distribution difference: we consider the concatenation of features 6-21, along with the concatenation of features 31-46, as two distributions, and calculate the Jensen–Shannon divergence between them.
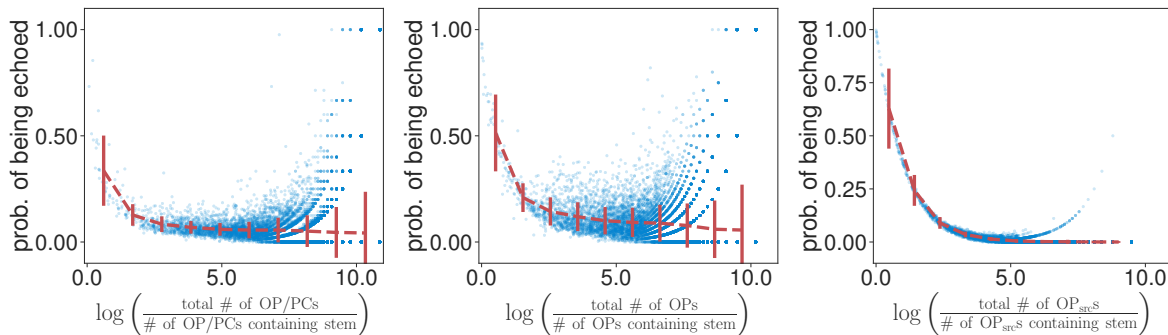
(a) Echoing probability between explananda and explanations.
(b) Echoing probability between OPs and their PCs.
(c) Echoing probability between OPs and other, randomly chosen OPs.

Figure 1: The U-shape exists both in Figure 1a and Figure 1b, but not in Figure 1c.

60. Stem dependency distribution difference: similarly, we consider the concatenation of features 22-24 (OP dependency labels), and the concatenation of features 47-49 (PC dependency labels), as two distributions, and calculate the Jensen–Shannon divergence between them.

61. OP length: the number of tokens in the OP.

62. PC length: the number of tokens in the PC.

63. Length difference: the absolute value of the difference between OP length and PC length.

64. Avg. word length difference: the difference between the average number of characters per token in the OP and the average number of characters per token in the PC.

65. OP/PC part–of–speech tag distribution difference: the Jensen–Shannon divergence between the part–of–speech tag distributions of the OP on the one hand, and the PC on the other.

66. Depth of the PC in the thread: since there can be many back–and–forth replies before a user awards a delta, we number each comment in a thread, starting at 0 for the OP, and incrementing for each new comment before the PC appears.

A.4 Word–level Prediction Task

For each non–LSTM classifier, we train 11 models: one full model, and forward and backward models for each of the five feature groups. To train, we fit on the training set and use the validation set for hyperparameter tuning.

For the random model, since the echo rate of the training set is 15%, we simply predict 1 with 15% probability, and 0 otherwise.

For logistic regression, we use the `lbfgs` solver. To tune hyperparameters, we perform an exhaustive grid search, with C taking values from $\{10^x : x \in \{-1, 0, 1, 2, 3, 4\}\}$, and the respective weights of the negative and positive classes taking values from $\{(x, 1 - x) : x \in \{0.25, 0.20, 0.15\}\}$.

We also train XGBoost models. Here, we use a learning rate of 0.1, 1000 estimator trees, and no subsampling. We perform an exhaustive grid search to tune hyperparameters, with the max tree depth equaling 5, 7, or 9, the minimum weight of a child equaling 3, 5, or 7, and the weight of a positive class instance equaling 3, 4, or 5.

Finally, we train two LSTM models, each with a single 300–dimensional hidden layer. Due to efficiency considerations, we eschewed a full search of the parameter space, but experimented with different values of dropout, learning rate, positive class weight, and batch size. We ultimately trained each model for five epochs with a batch size of 32 and a learning rate of 0.001, using the Adam optimizer (Kingma and Ba, 2015). We also weight positive instances four times more highly than negative instances.

A.5 Generating Explanations

We formulate an abstractive summarization task using an OP concatenated with the PC as a source, and the explanation as target. We train two models, one with the features described above, and one without. A shared vocabulary of 50k words is constructed from the training set by setting the maximum encoding length to 500 words. We set the maximum decoding length to 100. We use a pointer generator network with coverage for generating explanations, using a bidirectional LSTM

as an encoder and a unidirectional LSTM as a decoder. Both use a 256-dimensional hidden state. The parameters of this network are tuned using a validation set of five thousand instances. We constrain the batch size to 16 and train the network for 20k steps, using the parameters described in Table 4.

References

Center for Language and Information Research. 2016. ClearNLP Guidelines. https://github.com/clir/clearnlp-guidelines/blob/master/md/specifications/dependency_labels.md. [Online; accessed 21-May-2019].

Matthew Honnibal and Ines Montani. 2017. spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing.

Diederik P. Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *Proceedings of ICLR*.

Edward Loper and Steven Bird. 2002. NLTK: the natural language toolkit. *arXiv preprint cs/0205028*.

Sebastian Schuster and Christopher Manning. 2016. Enhanced English Universal Dependencies: an Improved Representation for Natural Language Understanding Tasks. *LREC 2016*.

Feature	all	content words	stopwords
Inverse document frequency	↓↓↓↓	↓↓↓↓	↓↓↓↓
Stem length	↓↓↓↓	↓↓↓↓	↓↓↓↓
Wordnet depth (min)	↓↓↓↓	↑↑↑↑	↑↑↑↑
Wordnet depth (max)	↓↓↓↓	↑↑↑↑	↑↑↑↑
Stem transfer probability	↑↑↑↑	↑↑↑↑	↑↑↑↑
OP ADP	↑↑↑↑	↑↑↑↑	↑↑↑↑
OP PRON	↑↑↑↑	↑↑↑↑	↑↑↑↑
OP X	↓↓↓↓	↑↑↑↑	↓↓↓↓
OP DET	↑↑↑↑	↑↑↑↑	↑↑↑↑
OP ADJ	↓↓↓↓	↓↓↓↓	↓↓↓↓
OP PROPN	↓↓↓↓	↓↓↓↓	↓↓↓↓
OP VERB	↓↓↓↓	↓↓	↑↑↑↑
OP PART	↑↑↑↑	↑↑↑↑	↑↑↑↑
OP CCONJ	↑↑↑↑	↑↑↑↑	↑↑↑↑
OP INTJ	↓↓↓↓	↑↑↑↑	↓↓↓↓
OP NOUN	↓↓↓↓	↓↓↓↓	↓↓↓↓
OP NUM	↓↓↓↓	↓↓	↓↓↓↓
OP ADV	↓↓↓↓	↑↑↑↑	↓↓↓↓
OP PUNCT	↓↓↓↓	↑↑↑↑	↓↓↓↓
OP SYM	↓↓↓↓	↑↑↑↑	↓↓↓↓
OP AUX	↓↓↓↓	↑↑↑↑	↓↓↓↓
OP subject	↑↑↑↑	↑↑↑↑	↑↑↑↑
OP object	↓↓↓↓	↑↑↑↑	↓↓↓↓
OP other	↑↑↑↑	↓↓↓↓	↑↑↑↑
OP term frequency	↑↑↑↑	↑↑↑↑	↑↑↑↑
OP normalized term frequency	↑↑↑↑	↑↑↑↑	↑↑↑↑
OP # of surface forms	↑↑↑↑	↑↑↑↑	↑↑↑↑
OP location	↑↑↑↑	↑↑↑↑	—
OP in quotes	↑↑↑↑	↑↑↑↑	↑↑↑↑
OP is entity	↑↑↑↑	↑↑↑↑	↑↑↑↑
PC ADP	↑↑↑↑	↓↓↓↓	↑↑↑↑
PC PRON	↑↑↑↑	↓↓↓↓	↑↑↑↑
PC X	↓↓↓↓	↓↓↓↓	↓↓↓↓
PC DET	↑↑↑↑	↓↓↓↓	↑↑↑↑
PC ADJ	↓↓↓↓	↑↑↑↑	↓↓↓↓
PC PROPN	↓↓↓↓	↓↓↓↓	↓↓↓↓
PC VERB	↑↑↑↑	↑↑↑↑	↑↑↑↑
PC PART	↓↓↓↓	↓↓↓↓	↑↑↑↑
PC CCONJ	↑↑↑↑	↓↓↓↓	↑↑↑↑
PC INTJ	↓↓↓↓	↓↓↓↓	↓↓↓↓
PC NOUN	↓↓↓↓	↑↑↑↑	↓↓↓↓
PC NUM	↓↓↓↓	↓↓↓↓	↓↓↓↓
PC ADV	—	↑↑↑↑	↓↓↓↓
PC PUNCT	↓↓↓↓	↓↓↓↓	↓↓↓↓
PC SYM	↓↓↓↓	↓↓↓↓	↓↓↓↓
PC AUX	↓↓↓↓	↓↓↓↓	↓↓↓↓
PC subject	↓↓↓↓	↓↓↓↓	↓↓↓↓
PC object	↓↓↓↓	↓↓↓↓	↓↓↓↓
PC other	↑↑↑↑	↑↑↑↑	↑↑↑↑
PC term frequency	↑↑↑↑	↑↑↑↑	↑↑↑↑
PC normalized term frequency	↑↑↑↑	↑↑↑↑	↑↑↑↑
PC # of surface forms	↑↑↑↑	↑↑↑↑	↑↑↑↑
PC location	↑	↑↑↑↑	↓↓↓↓
PC in quotes	↑↑↑↑	↑↑↑↑	↑↑↑↑
PC is entity	↑↑↑↑	↑↑↑↑	↑↑↑↑
In both OP and PC	↑↑↑↑	↑↑↑↑	↑↑↑↑
# of unique surface forms in OP	↓↓↓↓	↓↓↓↓	↓↓↓↓
# of unique surface forms in PC	↑↑↑↑	↑↑↑↑	↓↓↓↓
Stem POS distribution difference	↓↓↓↓	↓↓↓↓	↓↓↓↓
Stem dependency distribution difference	↓↓↓↓	↓↓↓↓	↓↓↓↓
OP length	↓↓↓↓	↓↓↓↓	↑↑↑↑
PC length	↑↑↑↑	↑↑↑↑	↑↑↑↑
Length difference	↓↓↓↓	↓↓↓↓	↑↑↑↑
Avg. word length difference	↓↓↓↓	↓↓↓↓	↓↓↓↓
OP/PC POS distribution difference	↓↓↓↓	↓↓↓↓	↓↓↓↓
Depth of the PC in the thread	↑↑↑↑	↑↑↑↑	↑↑↑↑

Table 2: Full testing results after Bonferroni correction.

Feature	Total Gain (%)
Inverse document frequency	16.97
Stem length	0.15
Wordnet depth (min)	0.12
Wordnet depth (max)	0.1
Stem transfer probability	46.7
OP ADP	0.02
OP PRON	0.1
OP X	0.01
OP DET	0.02
OP ADJ	0.01
OP PROPN	0.02
OP VERB	0.04
OP PART	0.01
OP CCONJ	0.0
OP INTJ	0.01
OP NOUN	0.04
OP NUM	0.01
OP ADV	0.15
OP PUNCT	0.01
OP SYM	0.0
OP AUX	0.0
OP subject	0.53
OP object	0.01
OP other	0.02
OP term frequency	3.23
OP normalized term frequency	0.26
OP # of surface forms	0.01
OP location	0.15
OP in quotes	0.01
OP is entity	0.02
PC ADP	0.02
PC PRON	0.09
PC X	0.81
PC DET	0.05
PC ADJ	0.01
PC PROPN	0.02
PC VERB	0.01
PC PART	0.02
PC CCONJ	0.13
PC INTJ	0.1
PC NOUN	0.04
PC NUM	0.70
PC ADV	0.02
PC PUNCT	0.03
PC SYM	0.2
PC AUX	0.0
PC subject	0.01
PC object	0.01
PC other	0.02
PC term frequency	3.33
PC normalized term frequency	2.92
PC # of surface forms	0.02
PC location	0.24
PC in quotes	0.04
PC is entity	0.02
In both OP and PC	4.88
# of unique surface forms in OP	0.01
# of unique surface forms in PC	0.03
Stem POS distribution difference	0.29
Stem dependency distribution difference	0.28
OP length	3.63
PC length	3.47
Length difference	2.59
Avg. word length difference	2.65
OP/PC POS distribution difference	3.15
Depth of the PC in the thread	1.4

Table 3: Feature importance for the full XGBoost model, as measured by total gain.

	Without features	With features
encoder type	brnn	brnn
glove vector dimension	300	300
rnn size	512	512
dropout	0.2	0.1
optim	adagrad	adam
learning rate	0.15	0.001
beam size	10	10

Table 4: Parameters tuned on validation dataset containing 5k instances.

Original Post:I keep seeing this point when people bitch about escort quests . But I 've been thinking about it and like , consider the alternatives : 1) The NPC moves at your walking speed . Clearly this is a terrible option . Nobody has ever willingly moved at their walking speed in a video game unless they were trying to finesse something or sneak . a walking speed escort quest would be terrible . The fact people even mention this point when talking about NPCs is insane . The actual complaint is " NPC is slower than my run speed " . If the NPC exclusively moved your walk speed it would be 10000 times worse . 2) The NPC moves at your run speed . This seems better at first ... but it means that you ca n't pull ahead of the NPC if you want to , or catch up to them if they ever get ahead of you because you stopped to do anything . They 're moving at 100 % of your max speed . Monsters up ahead ? That 's a fucking shame because you do n't have time to run up and pull aggro on them if the NPC is behind you and you are n't going to be able to intercept them in time at all if the NPC 's ahead because they 'll always get there first . 3) The NPC moves at your exact speed behind you following your pathfinding and dynamically navigating traps / moving parts to keep a uniform distance from you . This renders the escort quest pointless . This is a solution to a different problem (that escort quests are just terrible) . If having to escort and NPC does n't have any effect on your gameplay decisions they 're a pointless inclusion . The NPC is generally SUPPOSED to require your attention . After all , the only reason you care that they move slower than you is because you have to watch over them . In games where you do n't you just run ahead and let them be slow and it 's no problem . If you need to watch over them , then they need to act in a way where you ca n't just ignore them . Like , if you want to say escort quests are just terrible in general then I 'm on board . Escort quests suck . But if you have to have one with an NPC that has their own movement and pathfinding then they need to move close enough to your run speed that you are n't walking , but far enough from it that you can control your distance from them to some degree while ahead and can catch up when behind . Of all the options available for NPC movement speeds " about 75 - 85 % of PC run speed " is the best for escort missions both in terms of being least annoying for the player and most able to create the gameplay changes the devs want to create with escort quests . .

persuasive comment: I think the best solution is allowing the player to select a speed equal to the escort . The frustration does n't stem from having to move slower than you would without an escort . It 's that there 's no convenient way to move without running off and leaving them . In real life , it 's simple to adjust your movement speed to a slower person . It 's not about what pace you 're moving , it 's about being able to match pace .

Reference Explanation: Hmm ... maybe ? I was initially 100% convinced by *thinking* on it , I dunno . I feel that the annoying thing is just that they 're slow. Like , the fact you have to *run* laps on them when things are going *well* FEELS annoying but I *think* the ACTUAL annoying thing is just that it 's slow and because them *moving* slow causes that you fixate on it . *Running* at *pace* with them *would* be *equally* annoying because you 'd still know you COULD go faster ... That said you could be *right* and I 'm convinced enough that I *think* it 's worth a delta

Generated Explanation w/o features: This is a very good *point* . I had n't thought about it that *way* . Thank you for your time . I did n't *think* of it that *way*

Generated Explanation with features: I 'm going to give you a delta because you did n't change my *view* , but you 've convinced me that there is a difference between *escort* and *escort* ..

Table 5: Random generation from Open-NMT Pointer generator network with and without features. words like "escort" get copied from OP, but unable to construct a coherent, human-like sentence aggregating the context.

Original Post: Hi cmv , This post is not about whether or not abortion is morally permissible or ought to be legal . Rather , it 's a meta - view about the way the abortion debate is structured . Often , those on either side of the debate invoke the circumstances of the pregnancy to support their arguments . Speaking broadly , pro - choice advocates often point to sexual violence or lack of consent as a trump example . Pro - life advocates tend to argue that sex is a responsibility and that women who engage in casual sex are obligated to see a pregnancy through based on that decision . Logically , however , I ca n't see how the circumstances of a pregnancy hold bearing on whether an abortion is morally justifiable . Once a pregnancy has occurred , via any course of action , the moral quandary is the same - does the mother 's right to bodily autonomy take precedence over the fetus ' right to life ? Pick your favorite set of hypothetical circumstances , but at the end of the day the decision at hand is the same , and the logic that brings you to your conclusion ought to apply universally . While I understand the gut instinct to bring up rape and promiscuity when discussing this sensitive issue , I fail to see what bearing they hold to the core question , and believe they only serve to cloud and emotionally charge the debate . I 'm interested in well - structured arguments that connect the circumstances of pregnancy to the abortion debate . CMV !

persuasive comment: A major part of the abortion debate on the Pro - Life side is about legitimacy of the action . To someone that is Pro - Life abortion is the killing of a human being , but the killing of a human being is not by default murder . It is only murder if that killing is both unjustifiable and illegal . For those that want abortion to be classified as murder that means that they have to determine if the act is justifiable or not , and for many it is not an all or nothing proposition . Many who are Pro - Life consider abortion due to medical risk to the mother , severe disability of the child , or the child being the result of rape to be legitimate reasons to get an abortion and thus justifiable and in turn action that should be legal . It is the use of an abortion simply because you do not want a child , or because you forgot to use protection that they consider murder . This distinction in kind of human killing is also why killing in war , self defense , defense of another , and an actual pure accident not due to negligence are not crimes as well as why killing as an accident due to negligence is classified as manslaughter and not murder.

Reference Explanation: To *someone* that is *Pro - Life abortion* is the *killing* of a *human* being , but the *killing* of a *human* being is not by *default murder* . It is only *murder* if that *killing* is both *unjustifiable* and illegal. For those that *want abortion* to be *classified* as *murder* that *means* that they have to *determine* if the *act* is *justifiable* or not , and for *many* it is not an all or *nothing proposition* . You *get* a delta for this . The *distinction* between *ending* a *life* and *murder* is valid , and when operating with the latter definition , it makes sense that you 'd have to *consider* the extraneous factors . I may be limiting the *debate* too much by interpreting the *pro - life* position to be against *ending* a *life rather* than against *murder* .

Generated Explanation w/o features: "It is only about *whether* or not *abortion* is *morally permissible* or *ought* to be *legal*". *Rather* , it 's a *meta - view* about the *way* the *decision* at *hand* is the same , and the *killing* of a *human* , or the *child* being the *result* of *rape* to be *legitimate reasons* to *get* an *abortion* and *thus justifiable* and in *turn* that they have to *determine* if the *act* is *morally* . Once , *pro - choice advocates tend* to *argue* that *sex* is a *responsibility* and that *women* who *engage* in *casual sex* are

Generated Explanation with features: "It is only *murder* if that *killing* is both *unjustifiable* and *illegal* . For those that *want abortion* to be *classified* as *murder* that *means* that they have to *determine* if the *act* is *justifiable* or not , and for *many* it is not an all or *nothing proposition*". !delta. *Many* who are *Pro - Life consider abortion due* to *medical risk* that they *consider murder* . This *distinction* in *kind* of *human killing* is *also* why *killing* as an *accident due* to the *mother* , or the *child* being the *result* of *rape* to be *legitimate*

Table 6: Random generation from Open-NMT Pointer generator network with and without features.We can observe that generated explanations contain entire Quote from OP/PC indicating extractive summarization capabilities.

Original Post: When people use adblockers , they are hurting both consumers and producers . Adblockers take away the primary source of income for websites . Enough people use adblockers that this can seriously jeopardize the finances of a website . These sites include wikis , local newspapers , and many other valuable online resources . If the situation gets bad enough , it forces the producer to do one of 2 things . 1 . Shut down . OR 2 . Move to some sort of paid subscription service . Either way , the world just lost some valuable free information . This hurts the consumers . The benefits of adblockers are small compared to these consequences . Most people justify their use of adblockers by saying they want to avoid viruses / scams and/or intrusive / page - blocking / annoying ads If you are tech savvy enough to get an adblocker , you are probably tech savvy enough to understand what websites you should avoid . Plus you probably have an anti virus anyways . If you 're bothered by intrusive ads , just do n't visit the damn website . Shitty ads are the price you pay for going to some websites . If you are n't willing to pay that price , do n't go to those websites . That simple . That 's all I have to say I guess . i 've just seen too many good websites go down the drain because of this .	
persuasive comment: " Most people justify their use of adblockers by saying they want to avoid viruses / scams and/or intrusive / page - blocking / annoying ads " " If you are tech savvy enough to get an adblocker , you are probably tech savvy enough to understand what websites you should avoid . Plus you probably have an anti virus anyways . " Why avoid the website when you can neuter it with an adblocker ? It still has the content you were looking for after all . You 're also discounting the massive resource savings adblocking can cause . One university deployed an adblocker and saw their traffic go down 30 % . That 's huge , and that s just network resources , how many cpu cycles get wasted every second running poorly written javascript ads ? How much of your battery goes towards rendering ads ?	
Reference Explanation: That 's a *good* point about *resource wasting* . Not sure I 100% agree that that makes *adblockers* worth it but I think that 's a valid reason for *using one* . I 'll give you a delta	
Generated Explanation w/o features: That 's a *good* point . I had n't thought about it that *way* , but I 'll give you a delta for making me realize that it would be better. Thank you for changing my view	
Generated Explanation with features: I 'm *going* to give you a delta because I did n't really think of it in a *way* that makes sense to me . i 'm just *going* to give you a delta.	

Table 7: Random generation from Open-NMT Pointer generator network with and without features. Here we see both the generated examples fail to summarize resource wasting and generate a template explanation for changing the view.

Original Post: People 's main argument is that the poor will have to play money , but they would only have to pay very little because they make so little . It would make everyone feel that they are accomplishmisg something for the nation . Also I am also saying that the rich will also pay their fair ammounts . I forgot where but it was calcuated if everyone would pay a 24 % tax it would work out for America . Also I would argue for a flat tax because it completely eliminate the need for the Internal Revenue Service . Flat taxes are especially popular in the business and investment communities , where it is argued that because income from dividends , capital gains and distributions is untaxed , freeing up money that would have gone to taxes , investments and savings are thereby encouraged . But that is just my view , so feel free to change it . :)	
persuasive comment: I think it 's worth examining what you mean by fair . My read of your flat tax is that you think equal treatment gives equal / fair footing . Since you mentioned elsewhere that you think minimum wage should not be a permanent career choice , I assume you think people should start at minimum wage but take a better job as soon as possible . So , here are some practical questions : 1 . How is a person supposed to go on job interviews if they are already not making enough money to support them self ? 2 . How is a person supposed to spend money economically if they never have enough of it on a reliable basis to take advantage of sales , bulk purchases , payment plans , etc . ? 3 . How is a person supposed to got to school if they do n't have money to pay for their education ? And what responsible lender would loan someone that money knowing how little income they have available to make payments ? On the other side of the spectrum : 1 . How is a millionaire harmed by paying a higher marginal rate , given that he already has all of his basic need met , and then some ? 2 . What is the benefit to a millionaire putting his money into savings ? Does n't that just make less money available to the rest of the economy ? 3 . Is it ' fair ' to allow a millionaire to invest in and profit in a business , while paying those whose labor produces the profits for that business make less than a living wage ?	
Reference Explanation: *1* they *start* with a *minimum wage job* and *work* themselves up , that s how its always been *2* Plenty of ppl do nt have *enough money* right now , I am not *saying* that the system is perfect *3* There are plenty of *people* that *give loans* to students that do nt even have a *job* 4 He is *harmed* by losing *money* ? 5 They will use it to buy things thus boosting the econ . 6 Yea , because that s how they set it up and the gov . *allows* it to happen If anything boost the min . *wage* delta	
Generated Explanation w/o features: That 's a good point .Thanks for *changing* my view. delta	
Generated Explanation with features: That 's a good point . I was n't aware of it that way . I still *think* that it should be a *permanent career* , but that 's a good point .	

Table 8: Random generation from Open-NMT Pointer generator network with and without features.