DISAPERE

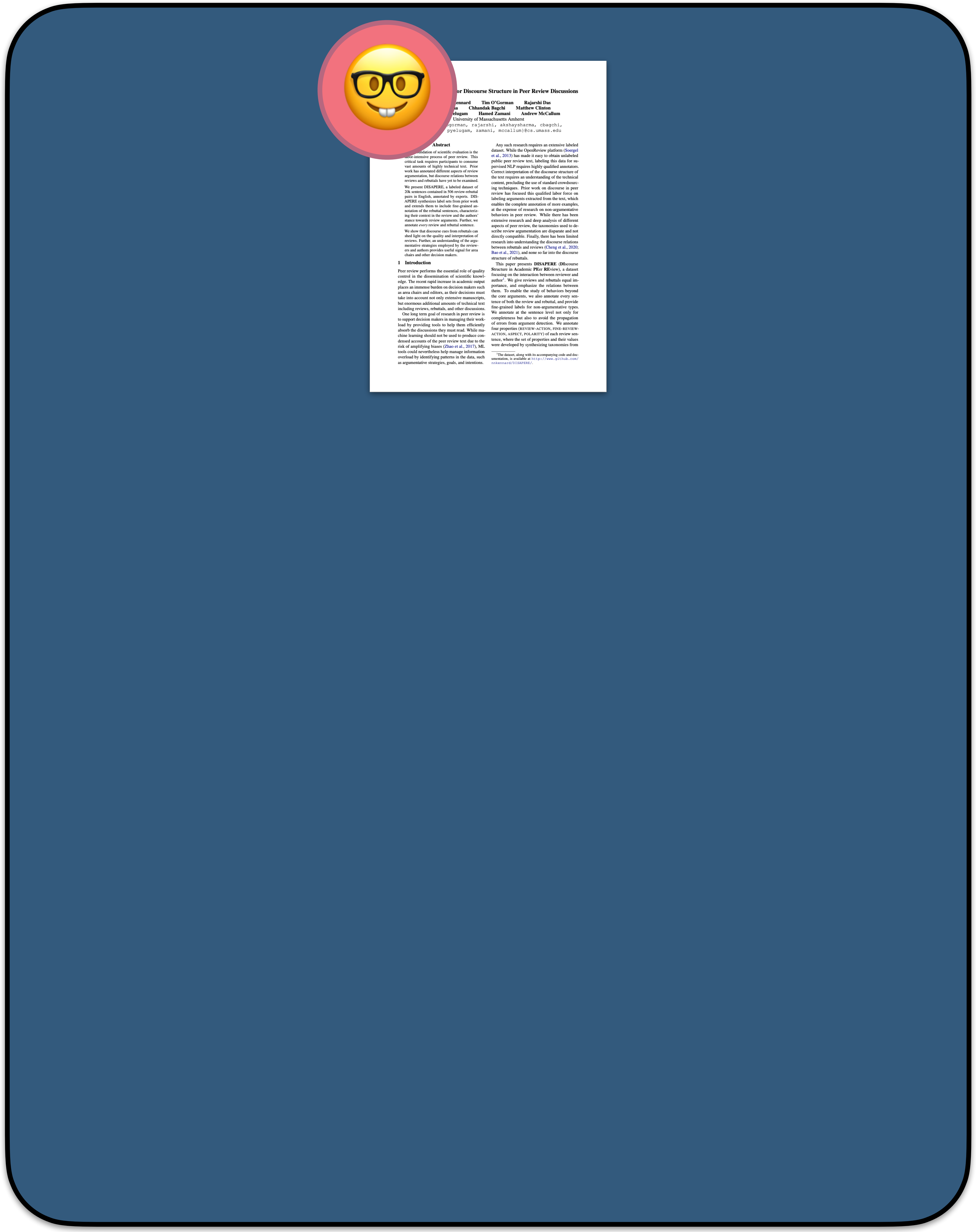
A Dataset for Discourse Structure in Peer Review Discussions

Neha Nayak Kennard Tim O'Gorman Rajarshi Das
Akshay Sharma Chhandak Bagchi Matthew Clinton Pranay Kumar Yelugam
Hamed Zamani Andrew McCallum

University of Massachusetts Amherst

Manuscript

Authors

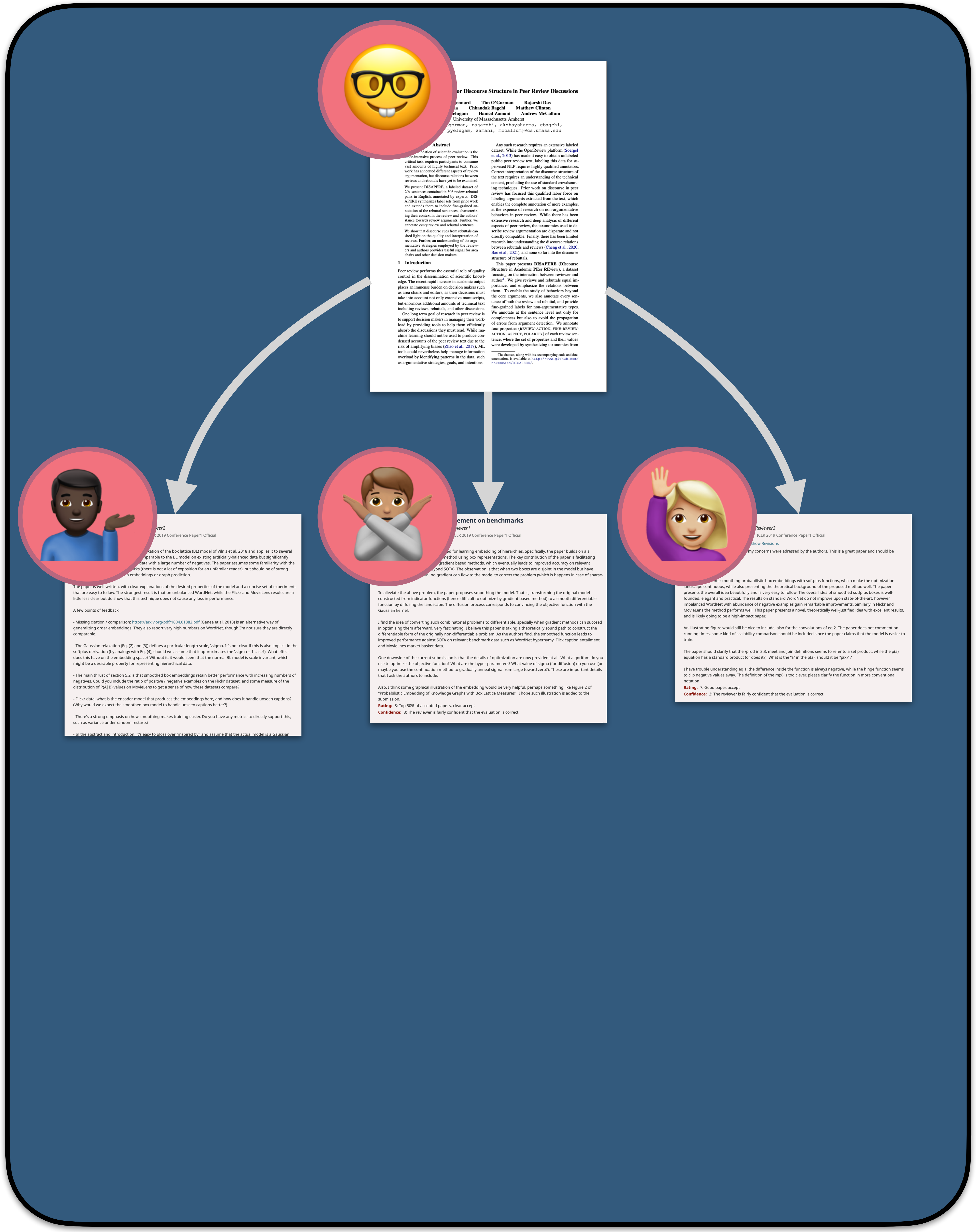


Manuscript

Authors

Reviews

Reviewers



Manuscript

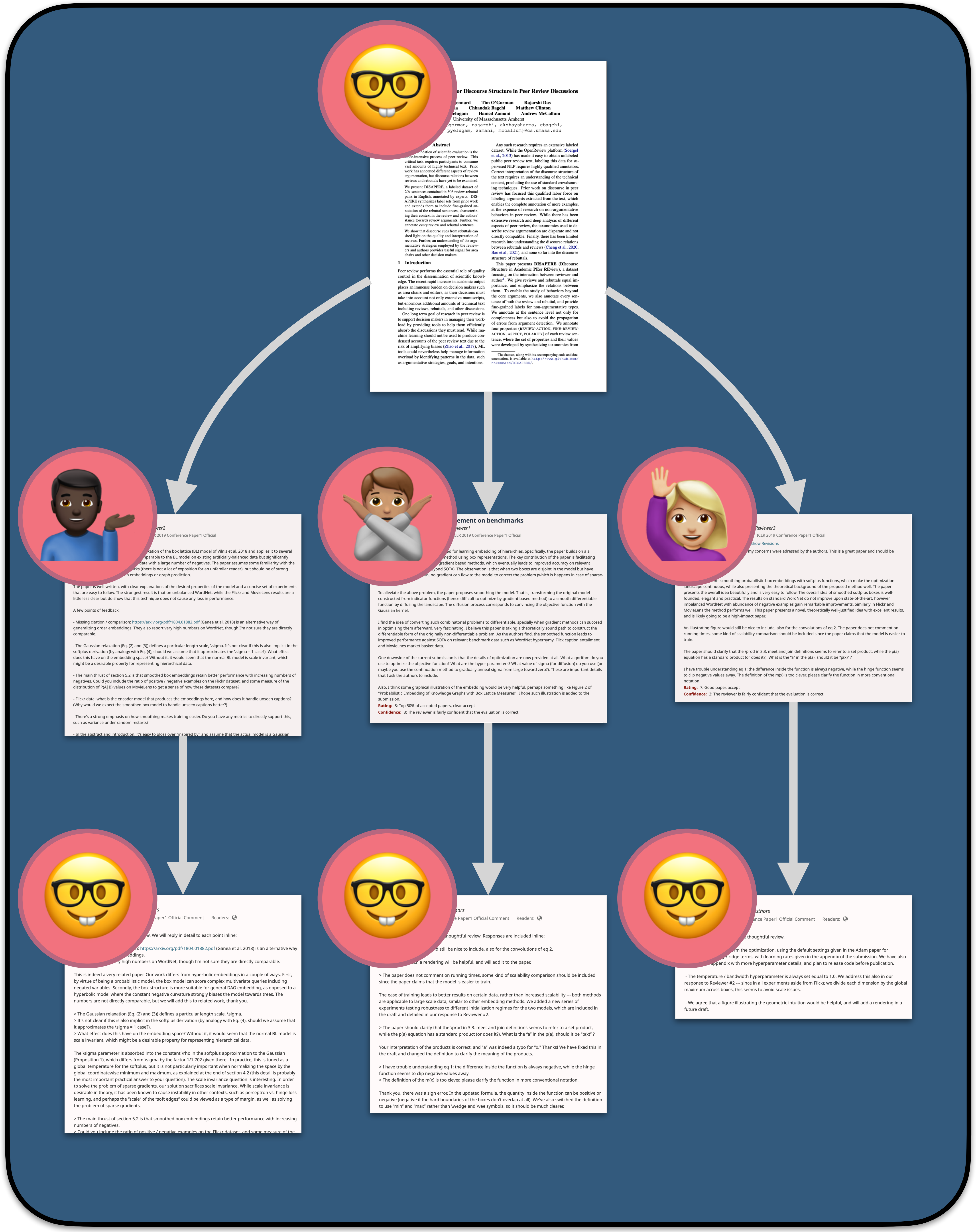
Authors

Reviews

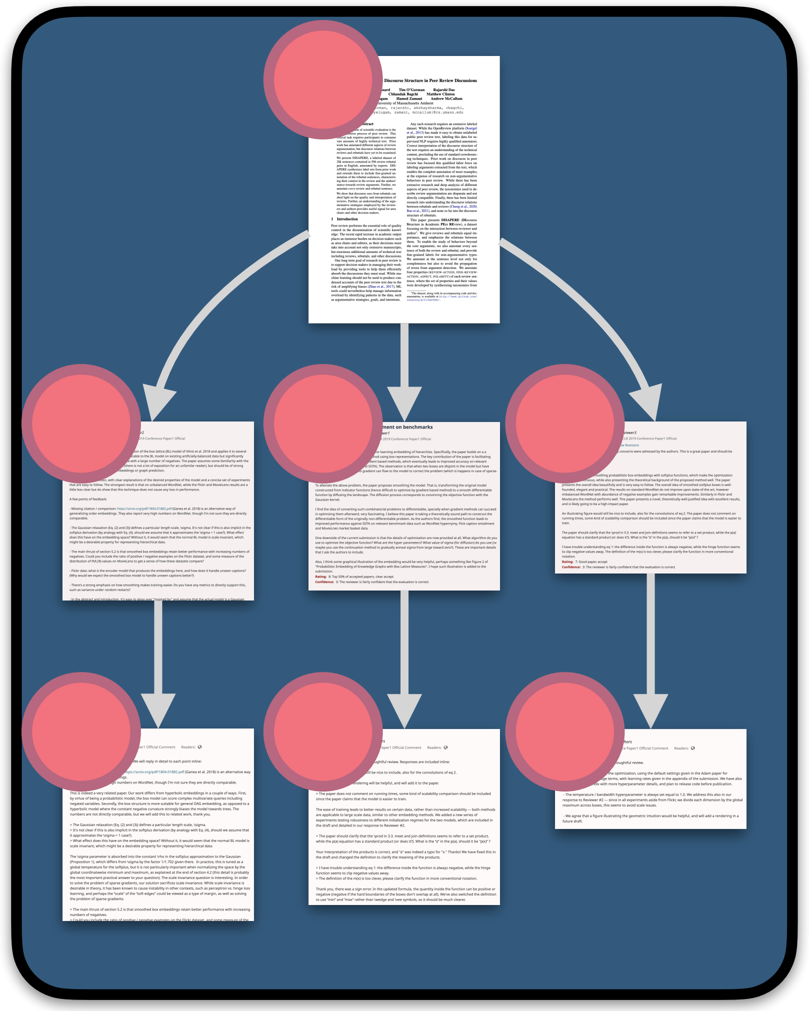
Reviewers

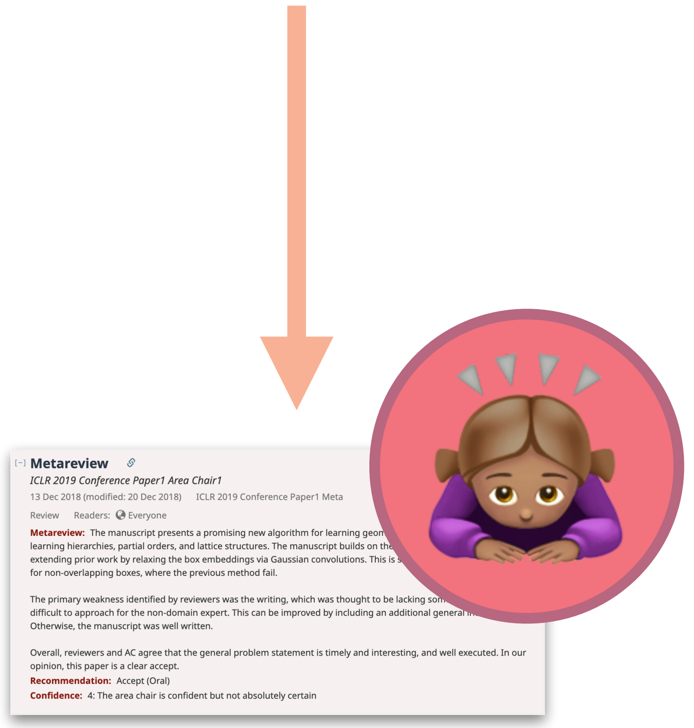
Rebuttals

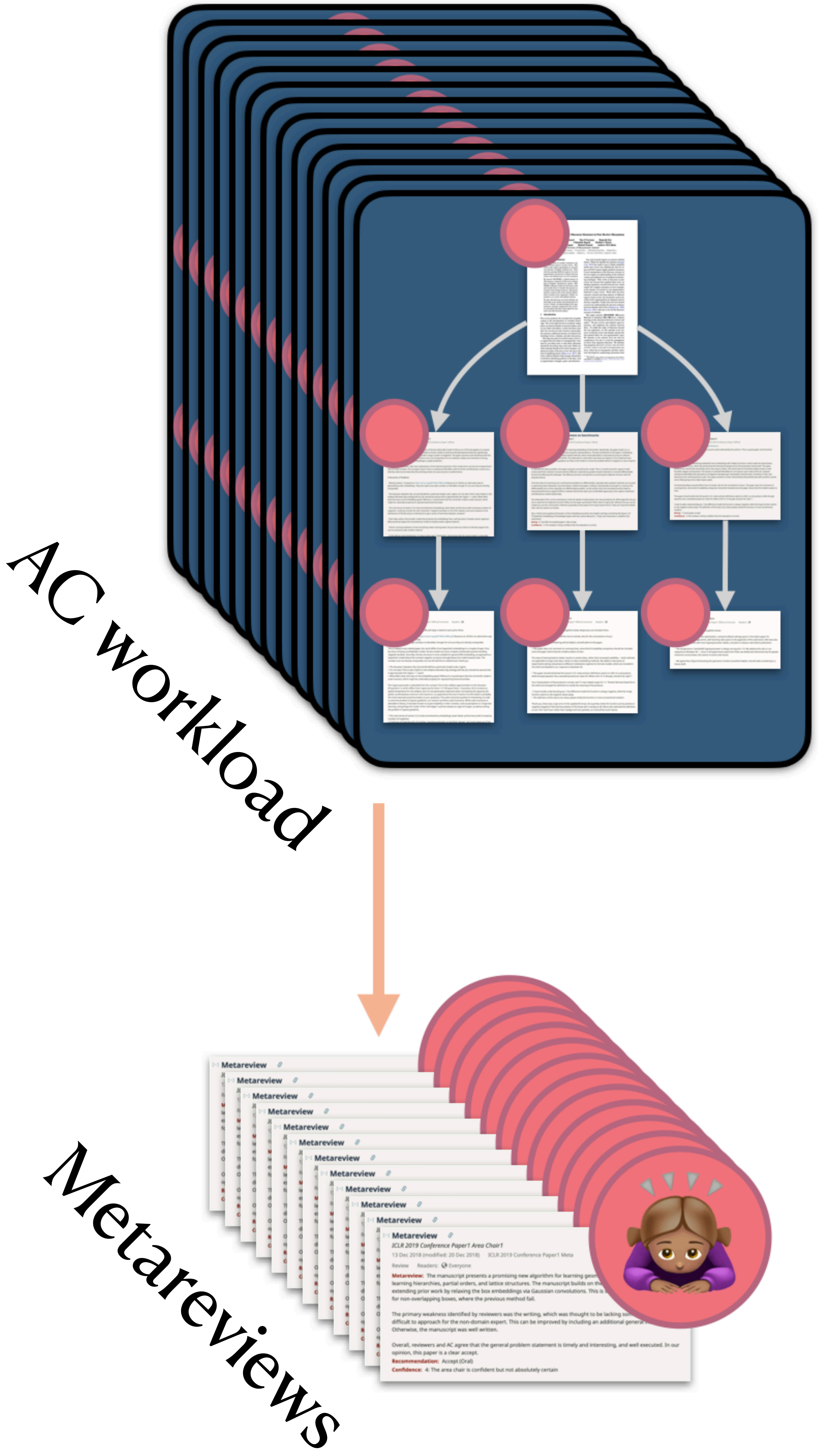
Authors



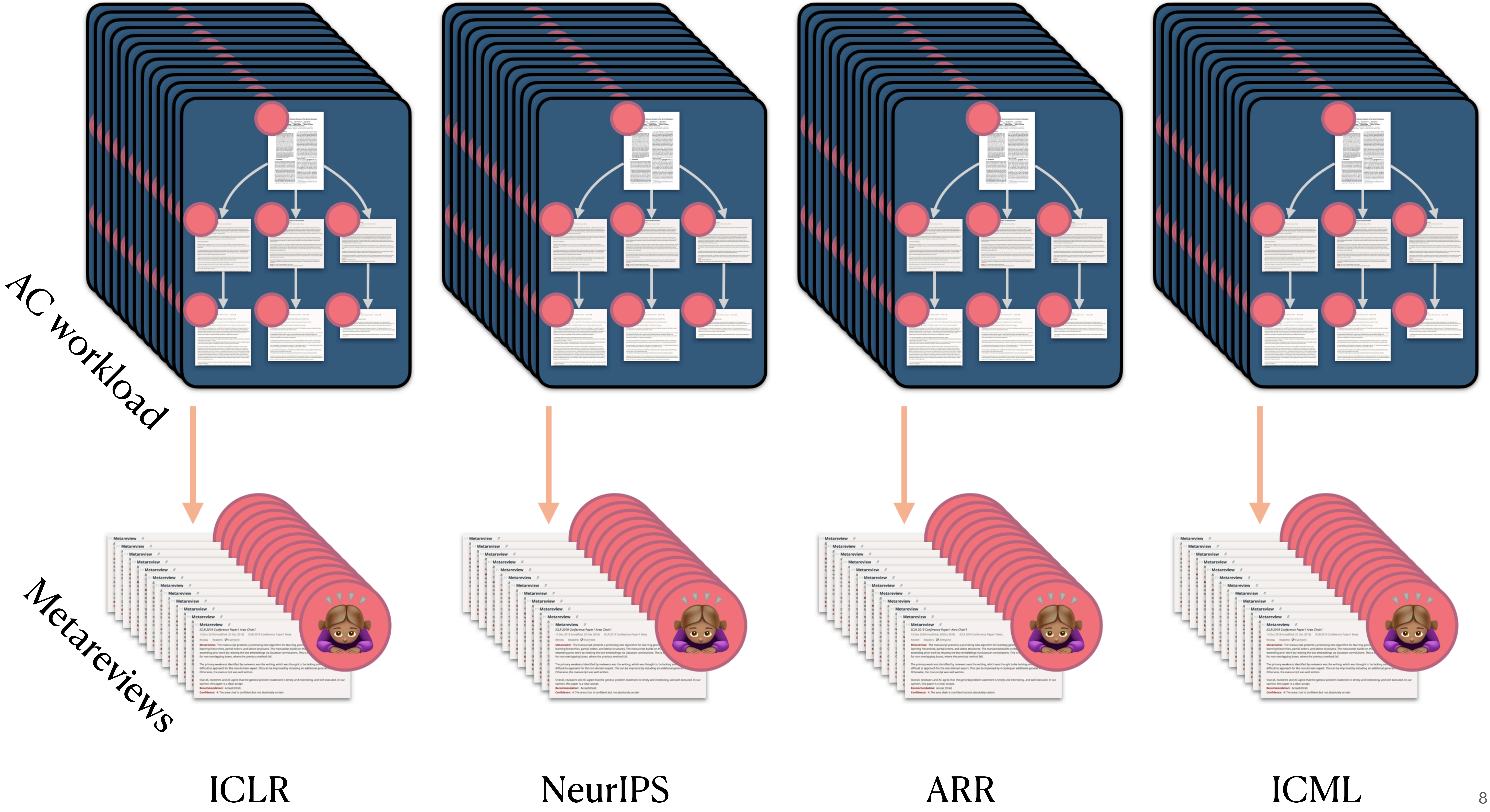
Manuscript
Authors
Reviews
Reviewers
Rebuttals
Authors
Metareview
Area chair (AC)




[−] Metareview
ICLR 2019 Conference Paper1 Area Chair1
13 Dec 2018 (modified: 20 Dec 2018) ICLR 2019 Conference Paper1 Meta
Review Readers: Everyone
Metareview: The manuscript presents a promising new algorithm for learning geom
learning hierarchies, partial orders, and lattice structures. The manuscript builds on the
extending prior work by relaxing the box embeddings via Gaussian convolutions. This is s
for non-overlapping boxes, where the previous method fail.
The primary weakness identified by reviewers was the writing, which was thought to be lacking som
difficult to approach for the non-domain expert. This can be improved by including an additional general
Otherwise, the manuscript was well written.
Overall, reviewers and AC agree that the general problem statement is timely and interesting, and well executed. In our opinion, this paper is a clear accept.
Recommendation: Accept (Oral)
Confidence: 4: The area chair is confident but not absolutely certain



ICLR


AC workload
Metareviews
ICLR
NeurIPS
ARR
ICML

Two important goals of peer review

Two important goals of peer review

Goal
Make correct decisions

Two important goals of peer review

Goal
Make correct decisions
Give constructive feedback

Two important goals of peer review

Goal	Evaluative metric
Make correct decisions	Score variance
Give constructive feedback	

Two important goals of peer review

Goal	Evaluative metric
Make correct decisions	Score variance
Give constructive feedback	?

DISAPERE dataset

DIscourse Structure in Academic PEer REview

DISAPERE dataset

DIscourse Structure in Academic PEer REview

5 classification tasks, 1 alignment task

506 review-rebuttal pairs

DISAPERE dataset

DIscourse Structure in Academic PEer REview

5 classification tasks, 1 alignment task

506 review-rebuttal pairs

Taken from ICLR 2019-2020 (OpenReview.net)

DISAPERE dataset

DIscourse Structure in Academic PEer REview

5 classification tasks, 1 alignment task

506 review-rebuttal pairs

Taken from ICLR 2019-2020 (OpenReview.net)

Over 21k sentences

Over 850 person-hours of expert annotation

Questions

Questions

Can we measure whether feedback is **constructive**?

Questions

Can we measure whether feedback is **constructive**?

Did the reviewer **try** to be helpful?

Questions

Can we measure whether feedback is **constructive**?

Did the reviewer **try** to be helpful?

Can we measure whether feedback is **applicable**?

Questions

Can we measure whether feedback is **constructive**?

Did the reviewer **try** to be helpful?

Can we measure whether feedback is **applicable**?

Did the reviewer **succeed** in being helpful?

Is the feedback constructive?

Sentence-level review labels

Argument Mining for Understanding Peer Reviews (Hua et al., NAACL 2019)
Can We Automate Scientific Reviewing? (Yuan et aal. arXiv 2021)
Argument Mining Driven Analysis of Peer-Reviews (Fromm et al., arXiv 2021)

Sentence-level review labels

Argumentative

Argument Mining for Understanding Peer Reviews (Hua et al., NAACL 2019)
Can We Automate Scientific Reviewing? (Yuan et aal. arXiv 2021)
Argument Mining Driven Analysis of Peer-Reviews (Fromm et al., arXiv 2021)

Sentence-level review labels

Evaluative *It's hard to evaluate whether the method would be useful.*

Argumentative

Argument Mining for Understanding Peer Reviews (Hua et al., NAACL 2019)
Can We Automate Scientific Reviewing? (Yuan et aal. arXiv 2021)
Argument Mining Driven Analysis of Peer-Reviews (Fromm et al., arXiv 2021)

Sentence-level review labels

	Evaluative	*It's hard to evaluate whether the method would be useful.*
Argumentative	Request	*How does the number of layers affect performance?*

Argument Mining for Understanding Peer Reviews (Hua et al., NAACL 2019)
Can We Automate Scientific Reviewing? (Yuan et aal. arXiv 2021)
Argument Mining Driven Analysis of Peer-Reviews (Fromm et al., arXiv 2021)

Sentence-level review labels

Argumentative	Evaluative	*It's hard to evaluate whether the method would be useful.*
	Request	*How does the number of layers affect performance?*
	Fact	*This is becoming a crowded space, with many such models.*

Argument Mining for Understanding Peer Reviews (Hua et al., NAACL 2019)
Can We Automate Scientific Reviewing? (Yuan et aal. arXiv 2021)
Argument Mining Driven Analysis of Peer-Reviews (Fromm et al., arXiv 2021)

[−] **Official Blind Review #3**

ICLR 2020 Conference Paper443 AnonReviewer3

24 Oct 2019 (modified: 05 Nov 2019) ICLR 2020 Conference Paper443 Official Review Readers: Everyone

Review: This paper presents a method for adapting a model that has been trained to perform one task, so that it can perform a new task (potentially without using any new training data at all—i.e., zero-shot learning). In some ways the presented work is a form of meta-learning or *meta-mapping* as the authors refer to it. The premise of the paper is very interesting and the overall problem is definitely of high interest and high potential impact.

I believe that the presentation of the proposed method can be significantly improved. The method description was a bit confusing and unclear to me. The experimental results presented were all done on small synthetic datasets and it's hard to evaluate whether the method is practically useful. Furthermore, no comparisons were provided to any baselines/alternative methods. For example, in Sections 4 and 5 I was hoping to see comparisons to methods like MAML. Also, I felt that the proposed approach in Section 5 is very similar to MAML intuitively. This makes a comparison with MAML even more desirable. Without any comparisons it's hard to tell how difficult the tasks under consideration are and what would amount to good performance on the held-out tasks.

In summary, I feel the paper tackles an interesting problem with an interesting approach, but the content could be organized much better. Also, this work would benefit significantly from a better experimental evaluation. For these reasons I lean towards rejecting this paper for now, but would love to see it refined for a future machine learning conference.

Also, the work by Platanios, et al. on contextual parameter generation is very relevant to this work as it tackles multi-task learning using HyperNetworks. It may be worth adding a short discussion/comparison to that work as it also considers zero-shot learning.

Minor comments:
- Capitalize: "section" -> "Section", "appendix" -> "Appendix", "fig." -> "Figure". Sometimes these are capitalized, but the use is inconsistent throughout the paper.
- "Hold-out" vs "held-out". Be consistent and use "held-out" throughout.

Rating: 3: Weak Reject
Experience Assessment: I have published one or two papers in this area.
Review Assessment: Thoroughness In Paper Reading: N/A
Review Assessment: Checking Correctness Of Derivations And Theory: N/A
Review Assessment: Checking Correctness Of Experiments: I assessed the sensibility of the experiments.

Reviewer recommendation:
3 (Reject)

[−] **Official Blind Review #3**

ICLR 2020 Conference Paper443 AnonReviewer3

24 Oct 2019 (modified: 05 Nov 2019) ICLR 2020 Conference Paper443 Official Review Readers: Everyone

Review: This paper presents a method for adapting a model that has been trained to perform one task, so that it can perform a new task (potentially without using any new training data at all—i.e., zero-shot learning). In some ways the presented work is a form of meta-learning or *meta-mapping* as the authors refer to it. The premise of the paper is very interesting and the overall problem is definitely of high interest and high potential impact.

I believe that the presentation of the proposed method can be significantly improved. The method description was a bit confusing and unclear to me. The experimental results presented were all done on small synthetic datasets and it's hard to evaluate whether the method is practically useful. Furthermore, no comparisons were provided to any baselines/alternative methods. For example, in Sections 4 and 5 I was hoping to see comparisons to methods like MAML. Also, I felt that the proposed approach in Section 5 is very similar to MAML intuitively. This makes a comparison with MAML even more desirable. Without any comparisons it's hard to tell how difficult the tasks under consideration are and what would amount to good performance on the held-out tasks.

In summary, I feel the paper tackles an interesting problem with an interesting approach, but the content could be organized much better. Also, this work would benefit significantly from a better experimental evaluation. For these reasons I lean towards rejecting this paper for now, but would love to see it refined for a future machine learning conference.

Also, the work by Platanios, et al. on contextual parameter generation is very relevant to this work as it tackles multi-task learning using HyperNetworks. It may be worth adding a short discussion/comparison to that work as it also considers zero-shot learning.

Minor comments:
- Capitalize: "section" -> "Section", "appendix" -> "Appendix", "fig." -> "Figure". Sometimes these are capitalized, but the use is inconsistent throughout the paper.
- "Hold-out" vs "held-out". Be consistent and use "held-out" throughout.

Rating: 3: Weak Reject
Experience Assessment: I have published one or two papers in this area.
Review Assessment: Thoroughness In Paper Reading: N/A
Review Assessment: Checking Correctness Of Derivations And Theory: N/A
Review Assessment: Checking Correctness Of Experiments: I assessed the sensibility of the experiments.

Reviewer recommendation: 3 (Reject)

This paper presents...

The premise is very interesting...

... whether the model is practically useful...

I was hoping to see...

... it's hard to tell how difficult the tasks...

... work would benefit from better evaluation...

... it may be worth adding a discussion of...

... capitalization...

Be consistent about...

Official Blind Review #3

ICLR 2020 Conference Paper443 AnonReviewer3

24 Oct 2019 (modified: 05 Nov 2019) ICLR 2020 Conference Paper443 Official Review Readers: Everyone

Review: This paper presents a method for adapting a model that has been trained to perform one task, so that it can perform a new task (potentially without using any new training data at all—i.e., zero-shot learning). In some ways the presented work is a form of meta-learning or *meta-mapping* as the authors refer to it. The premise of the paper is very interesting and the overall problem is definitely of high interest and high potential impact.

I believe that the presentation of the proposed method can be significantly improved. The method description was a bit confusing and unclear to me. The experimental results presented were all done on small synthetic datasets and it's hard to evaluate whether the method is practically useful. Furthermore, no comparisons were provided to any baselines/alternative methods. For example, in Sections 4 and 5 I was hoping to see comparisons to methods like MAML. Also, I felt that the proposed approach in Section 5 is very similar to MAML intuitively. This makes a comparison with MAML even more desirable. Without any comparisons it's hard to tell how difficult the tasks under consideration are and what would amount to good performance on the held-out tasks.

In summary, I feel the paper tackles an interesting problem with an interesting approach, but the content could be organized much better. Also, this work would benefit significantly from a better experimental evaluation. For these reasons I lean towards rejecting this paper for now, but would love to see it refined for a future machine learning conference.

Also, the work by Platanios, et al. on contextual parameter generation is very relevant to this work as it tackles multi-task learning using HyperNetworks. It may be worth adding a short discussion/comparison to that work as it also considers zero-shot learning.

Minor comments:
- Capitalize: "section" -> "Section", "appendix" -> "Appendix", "fig." -> "Figure". Sometimes these are capitalized, but the use is inconsistent throughout the paper.
- "Hold-out" vs "held-out". Be consistent and use "held-out" throughout.

Rating: 3: Weak Reject
Experience Assessment: I have published one or two papers in this area.
Review Assessment: Thoroughness In Paper Reading: N/A
Review Assessment: Checking Correctness Of Derivations And Theory: N/A
Review Assessment: Checking Correctness Of Experiments: I assessed the sensibility of the experiments.

Reviewer recommendation: 3 (Reject)

This paper presents...

The premise is very interesting...

... whether the model is practically useful...

I was hoping to see...

... it's hard to tell how difficult the tasks...

... work would benefit from better evaluation...

... it may be worth adding a discussion of...

... capitalization...

Be consistent about...

[−] **Official Blind Review #3**

ICLR 2020 Conference Paper443 AnonReviewer3

24 Oct 2019 (modified: 05 Nov 2019) ICLR 2020 Conference Paper443 Official Review Readers: Everyone

Review: This paper presents a method for adapting a model that has been trained to perform one task, so that it can perform a new task (potentially without using any new training data at all—i.e., zero-shot learning). In some ways the presented work is a form of meta-learning or *meta-mapping* as the authors refer to it. The premise of the paper is very interesting and the overall problem is definitely of high interest and high potential impact.

I believe that the presentation of the proposed method can be significantly improved. The method description was a bit confusing and unclear to me. The experimental results presented were all done on small synthetic datasets and it's hard to evaluate whether the method is practically useful. Furthermore, no comparisons were provided to any baselines/alternative methods. For example, in Sections 4 and 5 I was hoping to see comparisons to methods like MAML. Also, I felt that the proposed approach in Section 5 is very similar to MAML intuitively. This makes a comparison with MAML even more desirable. Without any comparisons it's hard to tell how difficult the tasks under consideration are and what would amount to good performance on the held-out tasks.

In summary, I feel the paper tackles an interesting problem with an interesting approach, but the content could be organized much better. Also, this work would benefit significantly from a better experimental evaluation. For these reasons I lean towards rejecting this paper for now, but would love to see it refined for a future machine learning conference.

Also, the work by Platanios, et al. on contextual parameter generation is very relevant to this work as it tackles multi-task learning using HyperNetworks. It may be worth adding a short discussion/comparison to that work as it also considers zero-shot learning.

Minor comments:
- Capitalize: "section" -> "Section", "appendix" -> "Appendix", "fig." -> "Figure". Sometimes these are capitalized, but the use is inconsistent throughout the paper.
- "Hold-out" vs "held-out". Be consistent and use "held-out" throughout.

Rating: 3: Weak Reject
Experience Assessment: I have published one or two papers in this area.
Review Assessment: Thoroughness In Paper Reading: N/A
Review Assessment: Checking Correctness Of Derivations And Theory: N/A
Review Assessment: Checking Correctness Of Experiments: I assessed the sensibility of the experiments.

Reviewer recommendation: 3 (Reject)

This paper presents...

The premise is very interesting...

... whether the model is practically useful...

I was hoping to see...

... it's hard to tell how difficult the tasks...

... work would benefit from better evaluation...

... it may be worth adding a discussion of...

... capitalization...

Be consistent about...

[−] **Official Blind Review #3**

ICLR 2020 Conference Paper443 AnonReviewer3

24 Oct 2019 (modified: 05 Nov 2019) ICLR 2020 Conference Paper443 Official Review Readers: Everyone

Review: This paper presents a method for adapting a model that has been trained to perform one task, so that it can perform a new task (potentially without using any new training data at all—i.e., zero-shot learning). In some ways the presented work is a form of meta-learning or *meta-mapping* as the authors refer to it. The premise of the paper is very interesting and the overall problem is definitely of high interest and high potential impact.

I believe that the presentation of the proposed method can be significantly improved. The method description was a bit confusing and unclear to me. The experimental results presented were all done on small synthetic datasets and it’s hard to evaluate whether the method is practically useful. Furthermore, no comparisons were provided to any baselines/alternative methods. For example, in Sections 4 and 5 I was hoping to see comparisons to methods like MAML. Also, I felt that the proposed approach in Section 5 is very similar to MAML intuitively. This makes a comparison with MAML even more desirable. Without any comparisons it’s hard to tell how difficult the tasks under consideration are and what would amount to good performance on the held-out tasks.

In summary, I feel the paper tackles an interesting problem with an interesting approach, but the content could be organized much better. Also, this work would benefit significantly from a better experimental evaluation. For these reasons I lean towards rejecting this paper for now, but would love to see it refined for a future machine learning conference.

Also, the work by Platanios, et al. on contextual parameter generation is very relevant to this work as it tackles multi-task learning using HyperNetworks. It may be worth adding a short discussion/comparison to that work as it also considers zero-shot learning.

Minor comments:
- Capitalize: “section” -> “Section”, “appendix” -> “Appendix”, “fig.” -> “Figure”. Sometimes these are capitalized, but the use is inconsistent throughout the paper.
- “Hold-out” vs “held-out”. Be consistent and use “held-out” throughout.

Rating: 3: Weak Reject
Experience Assessment: I have published one or two papers in this area.
Review Assessment: Thoroughness In Paper Reading: N/A
Review Assessment: Checking Correctness Of Derivations And Theory: N/A
Review Assessment: Checking Correctness Of Experiments: I assessed the sensibility of the experiments.

Reviewer recommendation: 3 (Reject)

This paper presents...

The premise is very interesting...

... whether the model is practically useful...

I was hoping to see...

... it’s hard to tell how difficult the tasks...

... work would benefit from better evaluation...

... it may be worth adding a discussion of...

... capitalization...

Be consistent about...

[−] **Official Blind Review #3**

ICLR 2020 Conference Paper443 AnonReviewer3

24 Oct 2019 (modified: 05 Nov 2019) ICLR 2020 Conference Paper443 Official Review Readers: Everyone

Review: This paper presents a method for adapting a model that has been trained to perform one task, so that it can perform a new task (potentially without using any new training data at all—i.e., zero-shot learning). In some ways the presented work is a form of meta-learning or *meta-mapping* as the authors refer to it. The premise of the paper is very interesting and the overall problem is definitely of high interest and high potential impact.

I believe that the presentation of the proposed method can be significantly improved. The method description was a bit confusing and unclear to me. The experimental results presented were all done on small synthetic datasets and it's hard to evaluate whether the method is practically useful. Furthermore, no comparisons were provided to any baselines/alternative methods. For example, in Sections 4 and 5 I was hoping to see comparisons to methods like MAML. Also, I felt that the proposed approach in Section 5 is very similar to MAML intuitively. This makes a comparison with MAML even more desirable. Without any comparisons it's hard to tell how difficult the tasks under consideration are and what would amount to good performance on the held-out tasks.

In summary, I feel the paper tackles an interesting problem with an interesting approach, but the content could be organized much better. Also, this work would benefit significantly from a better experimental evaluation. For these reasons I lean towards rejecting this paper for now, but would love to see it refined for a future machine learning conference.

Also, the work by Platanios, et al. on contextual parameter generation is very relevant to this work as it tackles multi-task learning using HyperNetworks. It may be worth adding a short discussion/comparison to that work as it also considers zero-shot learning.

Minor comments:
- Capitalize: "section" -> "Section", "appendix" -> "Appendix", "fig." -> "Figure". Sometimes these are capitalized, but the use is inconsistent throughout the paper.
- "Hold-out" vs "held-out". Be consistent and use "held-out" throughout.

Rating: 3: Weak Reject
Experience Assessment: I have published one or two papers in this area.
Review Assessment: Thoroughness In Paper Reading: N/A
Review Assessment: Checking Correctness Of Derivations And Theory: N/A
Review Assessment: Checking Correctness Of Experiments: I assessed the sensibility of the experiments.

Reviewer recommendation: 3 (Reject)

This paper presents...

The premise is very interesting...

... whether the model is practically useful...

I was hoping to see...

... it's hard to tell how difficult the tasks...

... work would benefit from better evaluation...

... it may be worth adding a discussion of...

... capitalization...

Be consistent about...

Official Blind Review #3

ICLR 2020 Conference Paper443 AnonReviewer3

24 Oct 2019 (modified: 05 Nov 2019) ICLR 2020 Conference Paper443 Official Review Readers: Everyone

Review: This paper presents a method for adapting a model that has been trained to perform one task, so that it can perform a new task (potentially without using any new training data at all—i.e., zero-shot learning). In some ways the presented work is a form of meta-learning or *meta-mapping* as the authors refer to it. The premise of the paper is very interesting and the overall problem is definitely of high interest and high potential impact.

I believe that the presentation of the proposed method can be significantly improved. The method description was a bit confusing and unclear to me. The experimental results presented were all done on small synthetic datasets and it’s hard to evaluate whether the method is practically useful. Furthermore, no comparisons were provided to any baselines/alternative methods. For example, in Sections 4 and 5 I was hoping to see comparisons to methods like MAML. Also, I felt that the proposed approach in Section 5 is very similar to MAML intuitively. This makes a comparison with MAML even more desirable. Without any comparisons it’s hard to tell how difficult the tasks under consideration are and what would amount to good performance on the held-out tasks.

In summary, I feel the paper tackles an interesting problem with an interesting approach, but the content could be organized much better. Also, this work would benefit significantly from a better experimental evaluation. For these reasons I lean towards rejecting this paper for now, but would love to see it refined for a future machine learning conference.

Also, the work by Platanios, et al. on contextual parameter generation is very relevant to this work as it tackles multi-task learning using HyperNetworks. It may be worth adding a short discussion/comparison to that work as it also considers zero-shot learning.

Minor comments:
- Capitalize: “section” -> “Section”, “appendix” -> “Appendix”, “fig.” -> “Figure”. Sometimes these are capitalized, but the use is inconsistent throughout the paper.
- “Hold-out” vs “held-out”. Be consistent and use “held-out” throughout.

Rating: 3: Weak Reject

Experience Assessment: I have published one or two papers in this area.

Review Assessment: Thoroughness In Paper Reading: N/A

Review Assessment: Checking Correctness Of Derivations And Theory: N/A

Review Assessment: Checking Correctness Of Experiments: I assessed the sensibility of the experiments.

Reviewer recommendation: 3 (Reject)

“I hope that an updated version will be accepted elsewhere.”

This paper presents...

The premise is very interesting...

... whether the model is practically useful...

I was hoping to see...

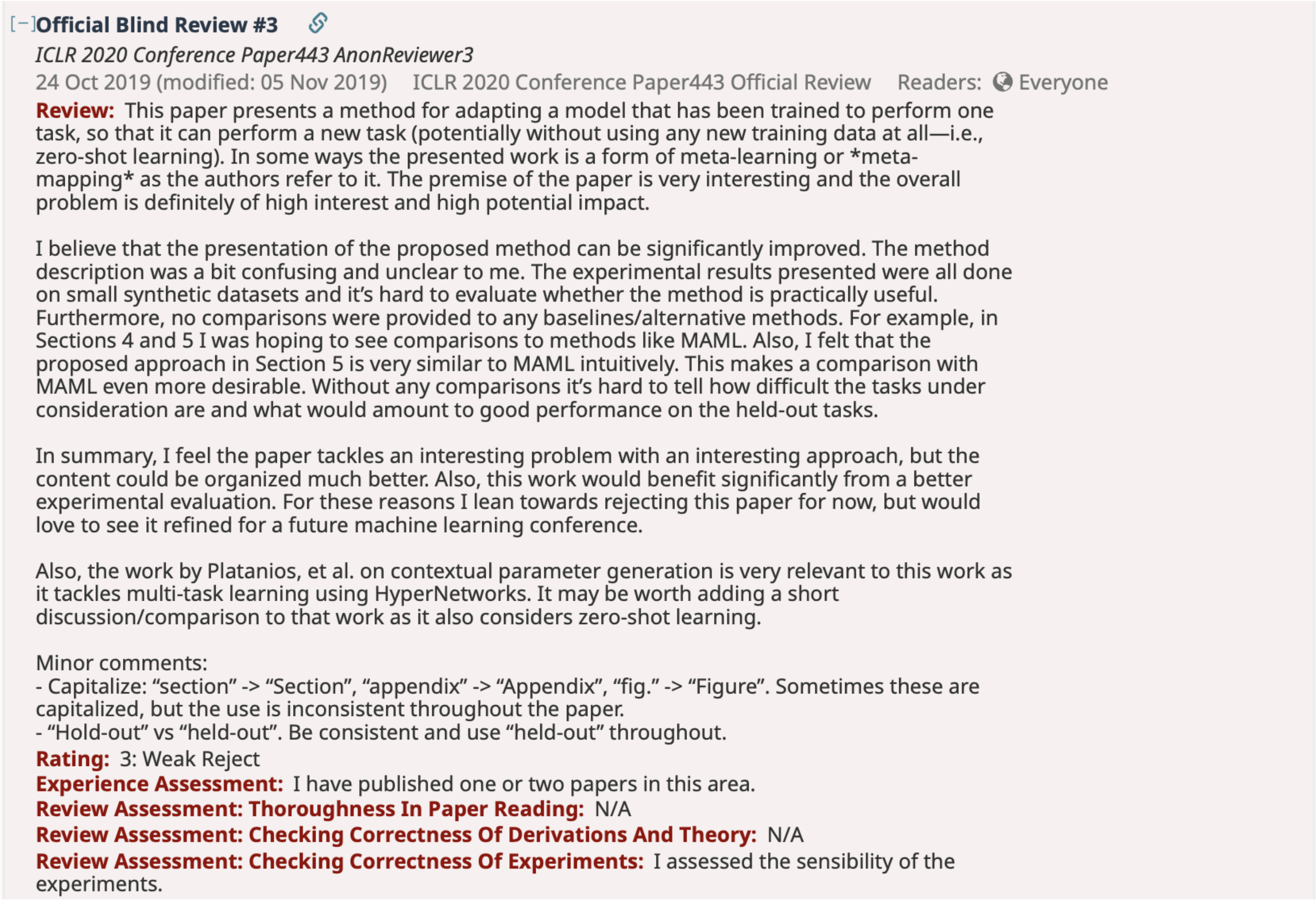
... it’s hard to tell how difficult the tasks...

... work would benefit from better evaluation...

... it may be worth adding a discussion of...

... capitalization...

Be consistent about...


[−]**Official Blind Review #3**

ICLR 2020 Conference Paper443 AnonReviewer3

24 Oct 2019 (modified: 05 Nov 2019) ICLR 2020 Conference Paper443 Official Review Readers: Everyone

Review: This paper presents a method for adapting a model that has been trained to perform one task, so that it can perform a new task (potentially without using any new training data at all—i.e., zero-shot learning). In some ways the presented work is a form of meta-learning or *meta-mapping* as the authors refer to it. The premise of the paper is very interesting and the overall problem is definitely of high interest and high potential impact.

I believe that the presentation of the proposed method can be significantly improved. The method description was a bit confusing and unclear to me. The experimental results presented were all done on small synthetic datasets and it's hard to evaluate whether the method is practically useful. Furthermore, no comparisons were provided to any baselines/alternative methods. For example, in Sections 4 and 5 I was hoping to see comparisons to methods like MAML. Also, I felt that the proposed approach in Section 5 is very similar to MAML intuitively. This makes a comparison with MAML even more desirable. Without any comparisons it's hard to tell how difficult the tasks under consideration are and what would amount to good performance on the held-out tasks.

In summary, I feel the paper tackles an interesting problem with an interesting approach, but the content could be organized much better. Also, this work would benefit significantly from a better experimental evaluation. For these reasons I lean towards rejecting this paper for now, but would love to see it refined for a future machine learning conference.

Also, the work by Platanios, et al. on contextual parameter generation is very relevant to this work as it tackles multi-task learning using HyperNetworks. It may be worth adding a short discussion/comparison to that work as it also considers zero-shot learning.

Minor comments:
- Capitalize: "section" -> "Section", "appendix" -> "Appendix", "fig." -> "Figure". Sometimes these are capitalized, but the use is inconsistent throughout the paper.
- "Hold-out" vs "held-out". Be consistent and use "held-out" throughout.

Rating: 3: Weak Reject

Experience Assessment: I have published one or two papers in this area.

Review Assessment: Thoroughness In Paper Reading: N/A

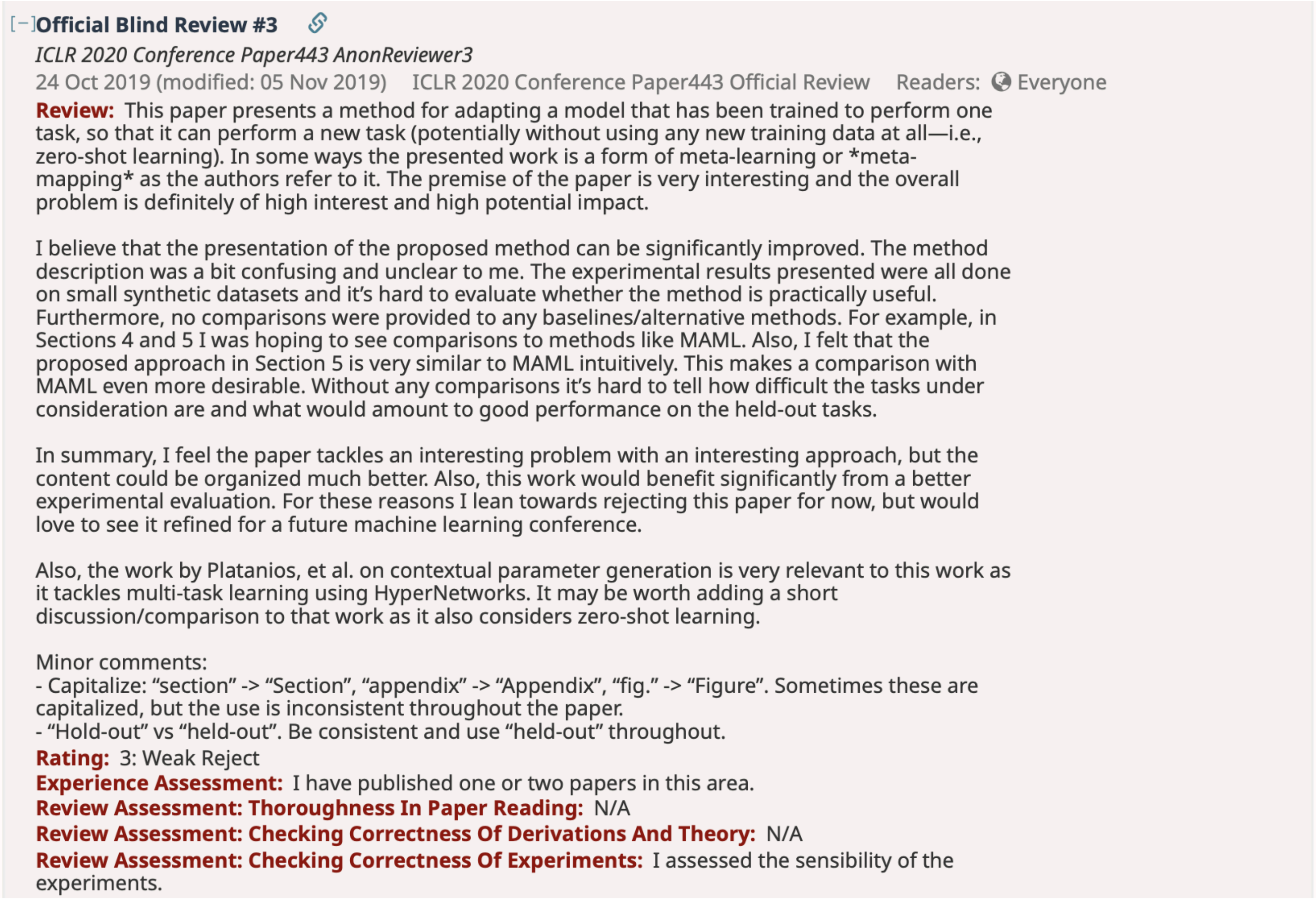
Review Assessment: Checking Correctness Of Derivations And Theory: N/A

Review Assessment: Checking Correctness Of Experiments: I assessed the sensibility of the experiments.

Reviewer recommendation: 3 (Reject)

"I hope that an updated version will be accepted elsewhere."

This paper presents...	Structuring
	Structuring
The premise is very interesting...	Evaluative
	Request
	Evaluative
... whether the model is practically useful...	Evaluative
	Evaluative
I was hoping to see...	Request
	Request
... it's hard to tell how difficult the tasks...	Evaluative
	Evaluative
... work would benefit from better evaluation...	Request
	Social
	Fact
... it may be worth adding a discussion of...	Request
	Structuring
	Request
	Evaluative
... capitalization...	Request
Be consistent about...	Request



[−] **Official Blind Review #3**

ICLR 2020 Conference Paper443 AnonReviewer3

24 Oct 2019 (modified: 05 Nov 2019) ICLR 2020 Conference Paper443 Official Review Readers: Everyone

Review: This paper presents a method for adapting a model that has been trained to perform one task, so that it can perform a new task (potentially without using any new training data at all—i.e., zero-shot learning). In some ways the presented work is a form of meta-learning or *meta-mapping* as the authors refer to it. The premise of the paper is very interesting and the overall problem is definitely of high interest and high potential impact.

I believe that the presentation of the proposed method can be significantly improved. The method description was a bit confusing and unclear to me. The experimental results presented were all done on small synthetic datasets and it's hard to evaluate whether the method is practically useful. Furthermore, no comparisons were provided to any baselines/alternative methods. For example, in Sections 4 and 5 I was hoping to see comparisons to methods like MAML. Also, I felt that the proposed approach in Section 5 is very similar to MAML intuitively. This makes a comparison with MAML even more desirable. Without any comparisons it's hard to tell how difficult the tasks under consideration are and what would amount to good performance on the held-out tasks.

In summary, I feel the paper tackles an interesting problem with an interesting approach, but the content could be organized much better. Also, this work would benefit significantly from a better experimental evaluation. For these reasons I lean towards rejecting this paper for now, but would love to see it refined for a future machine learning conference.

Also, the work by Platanios, et al. on contextual parameter generation is very relevant to this work as it tackles multi-task learning using HyperNetworks. It may be worth adding a short discussion/comparison to that work as it also considers zero-shot learning.

Minor comments:
- Capitalize: "section" -> "Section", "appendix" -> "Appendix", "fig." -> "Figure". Sometimes these are capitalized, but the use is inconsistent throughout the paper.
- "Hold-out" vs "held-out". Be consistent and use "held-out" throughout.

Rating: 3: Weak Reject
Experience Assessment: I have published one or two papers in this area.
Review Assessment: Thoroughness In Paper Reading: N/A
Review Assessment: Checking Correctness Of Derivations And Theory: N/A
Review Assessment: Checking Correctness Of Experiments: I assessed the sensibility of the experiments.

Reviewer recommendation: 3 (Reject)

"I hope that an updated version will be accepted elsewhere."

8/21 sentences are requests

This paper presents...	Structuring
	Structuring
The premise is very interesting...	Evaluative
	Request
	Evaluative
... whether the model is practically useful...	Evaluative
	Evaluative
I was hoping to see...	Request
	Request
... it's hard to tell how difficult the tasks...	Evaluative
	Evaluative
... work would benefit from better evaluation...	Request
	Social
	Fact
... it may be worth adding a discussion of...	Request
	Structuring
	Request
	Evaluative
... capitalization...	Request
Be consistent about...	Request

Official Blind Review #2

ICLR 2020 Conference Paper1166 AnonReviewer2

23 Oct 2019 (modified: 05 Nov 2019) ICLR 2020 Conference Paper1166 Official Review Readers: Everyone

Review: Summary:
This paper describes a contextual encoding scheme for reconstruction of 3D pointclouds from 2D images. An encoder outputs the parameters of a hierarchy of reconstruction networks that can be applied in succession to map random samples on a unit sphere to the surface of the reconstructed shape.

Strengths:
The author's model was quite novel in my opinion. Deep 2D->3D is becoming a crowded space and there are many other models that encode image inputs, and many others that perform recursive or composition-based decoding. However, the particular link here was interesting, and I appreciate the small number of parameters resulting in solid reconstruction performance. While most related work was covered well, I believe the authors could have a more up-to-date list of recent work that reconstructs triangle-mesh representations from images [A-C] (especially since several of these methods has an architecture that involves encoding and subsequent compositional refinement).

Some of the reconstructions shown in this paper are quite impressive, and the quantitative results show outperforming 2 recent methods. I did appreciate also the novel path-based evaluation of shape accuracy in the Appendix, although it would have been helpful to see more discussion of this in the main paper.

Areas for improvement:
I found that the core technical description was quite brief and would have benefited from simply more detail and space. You have argued that your method is sensible to try (cog. sci motivations), and shown that one instance works, but what can we expect in a more mathematical or general sense? Can any sizes of encoder and mapping network fit together? How does the number of mapping layers effect performance? Won't we eventually expect vanishing/exploding gradients with particular activation and can one address this in some way?

I note that recent papers in this field tend to perform significantly more extensive experimental evaluation, typically selecting a wider range of competitors and using a number of more standardized metrics including IOU, F1 score and CD and typically repeating these at a variety of resolutions or on additional datasets or category splits etc.

Decision:
Weak reject because the idea is quite interesting, but I believe a more thorough explanation and expanded experimental comparison would be of great help to ensure the community can appreciate this work.

Additional citations suggested:

[A] Pixel2Mesh: Generating 3D Mesh Models from Single RGB Images. Wang, Zhang, Li, Fu, Liu and Jiang. ECCV 2018.
[B] MeshCNN: A Network with an Edge. Hanocka, Hertz, Fish, Giryes, Fleishman and Cohen-Or. SIGGRAPH 2019.
[C] GEOMetrics: Exploiting Structure for Graph-Encoded Objects. Smith, Fujimoto, Romero and Meger. ICML 2019.

Rating: 3: Weak Reject
Experience Assessment: I have published in this field for several years.
Review Assessment: Thoroughness In Paper Reading: I read the paper thoroughly.
Review Assessment: Checking Correctness Of Derivations And Theory: I carefully checked the derivations and theory.
Review Assessment: Checking Correctness Of Experiments: I carefully checked the experiments.

Reviewer recommendation:
3 (Reject)

Official Blind Review #2

ICLR 2020 Conference Paper1166 AnonReviewer2

23 Oct 2019 (modified: 05 Nov 2019) ICLR 2020 Conference Paper1166 Official Review Readers: Everyone

Review: Summary:
This paper describes a contextual encoding scheme for reconstruction of 3D pointclouds from 2D images. An encoder outputs the parameters of a hierarchy of reconstruction networks that can be applied in succession to map random samples on a unit sphere to the surface of the reconstructed shape.

Strengths:
The author's model was quite novel in my opinion. Deep 2D->3D is becoming a crowded space and there are many other models that encode image inputs, and many others that perform recursive or composition-based decoding. However, the particular link here was interesting, and I appreciate the small number of parameters resulting in solid reconstruction performance. While most related work was covered well, I believe the authors could have a more up-to-date list of recent work that reconstructs triangle-mesh representations from images [A-C] (especially since several of these methods has an architecture that involves encoding and subsequent compositional refinement).

Some of the reconstructions shown in this paper are quite impressive, and the quantitative results show outperforming 2 recent methods. I did appreciate also the novel path-based evaluation of shape accuracy in the Appendix, although it would have been helpful to see more discussion of this in the main paper.

Areas for improvement:
I found that the core technical description was quite brief and would have benefited from simply more detail and space. You have argued that your method is sensible to try (cog. sci motivations), and shown that one instance works, but what can we expect in a more mathematical or general sense? Can any sizes of encoder and mapping network fit together? How does the number of mapping layers effect performance? Won't we eventually expect vanishing/exploding gradients with particular activation and can one address this in some way?

I note that recent papers in this field tend to perform significantly more extensive experimental evaluation, typically selecting a wider range of competitors and using a number of more standardized metrics including IOU, F1 score and CD and typically repeating these at a variety of resolutions or on additional datasets or category splits etc.

Decision:
Weak reject because the idea is quite interesting, but I believe a more thorough explanation and expanded experimental comparison would be of great help to ensure the community can appreciate this work.

Additional citations suggested:

[A] Pixel2Mesh: Generating 3D Mesh Models from Single RGB Images. Wang, Zhang, Li, Fu, Liu and Jiang. ECCV 2018.
[B] MeshCNN: A Network with an Edge. Hanocka, Hertz, Fish, Giryes, Fleishman and Cohen-Or. SIGGRAPH 2019.
[C] GEOMetrics: Exploiting Structure for Graph-Encoded Objects. Smith, Fujimoto, Romero and Meger. ICML 2019.

Rating: 3: Weak Reject
Experience Assessment: I have published in this field for several years.
Review Assessment: Thoroughness In Paper Reading: I read the paper thoroughly.
Review Assessment: Checking Correctness Of Derivations And Theory: I carefully checked the derivations and theory.
Review Assessment: Checking Correctness Of Experiments: I carefully checked the experiments.

Reviewer recommendation: 3 (Reject)

7/20 sentences are requests

Structuring
Structuring
Structuring
Structuring
Evaluative
Fact
Evaluative
Request
Evaluative
Evaluative
Structuring
Request
Request
Fact
Structuring
Request
Structuring
Request
Request
Request

[–] **Official Blind Review #2**

ICLR 2020 Conference Paper1166 AnonReviewer2

23 Oct 2019 (modified: 05 Nov 2019) ICLR 2020 Conference Paper1166 Official Review Readers: Everyone

Review: Summary:
This paper describes a contextual encoding scheme for reconstruction of 3D pointclouds from 2D images. An encoder outputs the parameters of a hierarchy of reconstruction networks that can be applied in succession to map random samples on a unit sphere to the surface of the reconstructed shape.

Strengths:
The author's model was quite novel in my opinion. Deep 2D->3D is becoming a crowded space and there are many other models that encode image inputs, and many others that perform recursive or composition-based decoding. However, the particular link here was interesting, and I appreciate the small number of parameters resulting in solid reconstruction performance. While most related work was covered well, I believe the authors could have a more up-to-date list of recent work that reconstructs triangle-mesh representations from images [A-C] (especially since several of these methods has an architecture that involves encoding and subsequent compositional refinement).

Some of the reconstructions shown in this paper are quite impressive, and the quantitative results show outperforming 2 recent methods. I did appreciate also the novel path-based evaluation of shape accuracy in the Appendix, although it would have been helpful to see more discussion of this in the main paper.

Areas for improvement:
I found that the core technical description was quite brief and would have benefited from simply more detail and space. You have argued that your method is sensible to try (cog. sci motivations), and shown that one instance works, but what can we expect in a more mathematical or general sense? Can any sizes of encoder and mapping network fit together? How does the number of mapping layers effect performance? Won't we eventually expect vanishing/exploding gradients with particular activation and can one address this in some way?

I note that recent papers in this field tend to perform significantly more extensive experimental evaluation, typically selecting a wider range of competitors and using a number of more standardized metrics including IOU, F1 score and CD and typically repeating these at a variety of resolutions or on additional datasets or category splits etc.

Decision:
Weak reject because the idea is quite interesting, but I believe a more thorough explanation and expanded experimental comparison would be of great help to ensure the community can appreciate this work.

Additional citations suggested:

[A] Pixel2Mesh: Generating 3D Mesh Models from Single RGB Images. Wang, Zhang, Li, Fu, Liu and Jiang. ECCV 2018.
[B] MeshCNN: A Network with an Edge. Hanocka, Hertz, Fish, Giryes, Fleishman and Cohen-Or. SIGGRAPH 2019.
[C] GEOMetrics: Exploiting Structure for Graph-Encoded Objects. Smith, Fujimoto, Romero and Meger. ICML 2019.

Rating: 3: Weak Reject
Experience Assessment: I have published in this field for several years.
Review Assessment: Thoroughness In Paper Reading: I read the paper thoroughly.
Review Assessment: Checking Correctness Of Derivations And Theory: I carefully checked the derivations and theory.
Review Assessment: Checking Correctness Of Experiments: I carefully checked the experiments.

This paper describes...	Structuring
...	Structuring
...	Structuring
...	Structuring
... was quite novel ...	Evaluative
...	Fact
...	Evaluative
... could highlight more recent work...	Request
...	Evaluative
I did appreciate...	Evaluative
...	Structuring
... benefit from more detail...	Request
...	Request
...	Fact
...	Structuring
...	Request
Additional citations suggested:	Structuring
[1]	Request
[2]	Request
[3]	Request

Reviewer recommendation: 3 (Reject)

7/20 sentences are requests

Official Blind Review #2

ICLR 2020 Conference Paper1166 AnonReviewer2

23 Oct 2019 (modified: 05 Nov 2019) ICLR 2020 Conference Paper1166 Official Review Readers: Everyone

Review: Summary:
This paper describes a contextual encoding scheme for reconstruction of 3D pointclouds from 2D images. An encoder outputs the parameters of a hierarchy of reconstruction networks that can be applied in succession to map random samples on a unit sphere to the surface of the reconstructed shape.

Strengths:
The author's model was quite novel in my opinion. Deep 2D->3D is becoming a crowded space and there are many other models that encode image inputs, and many others that perform recursive or composition-based decoding. However, the particular link here was interesting, and I appreciate the small number of parameters resulting in solid reconstruction performance. While most related work was covered well, I believe the authors could have a more up-to-date list of recent work that reconstructs triangle-mesh representations from images [A-C] (especially since several of these methods has an architecture that involves encoding and subsequent compositional refinement).

Some of the reconstructions shown in this paper are quite impressive, and the quantitative results show outperforming 2 recent methods. I did appreciate also the novel path-based evaluation of shape accuracy in the Appendix, although it would have been helpful to see more discussion of this in the main paper.

Areas for improvement:
I found that the core technical description was quite brief and would have benefited from simply more detail and space. You have argued that your method is sensible to try (cog. sci motivations), and shown that one instance works, but what can we expect in a more mathematical or general sense? Can any sizes of encoder and mapping network fit together? How does the number of mapping layers effect performance? Won't we eventually expect vanishing/exploding gradients with particular activation and can one address this in some way?

I note that recent papers in this field tend to perform significantly more extensive experimental evaluation, typically selecting a wider range of competitors and using a number of more standardized metrics including IOU, F1 score and CD and typically repeating these at a variety of resolutions or on additional datasets or category splits etc.

Decision:
Weak reject because the idea is quite interesting, but I believe a more thorough explanation and expanded experimental comparison would be of great help to ensure the community can appreciate this work.

Additional citations suggested:

[A] Pixel2Mesh: Generating 3D Mesh Models from Single RGB Images. Wang, Zhang, Li, Fu, Liu and Jiang. ECCV 2018.
[B] MeshCNN: A Network with an Edge. Hanocka, Hertz, Fish, Giryes, Fleishman and Cohen-Or. SIGGRAPH 2019.
[C] GEOMetrics: Exploiting Structure for Graph-Encoded Objects. Smith, Fujimoto, Romero and Meger. ICML 2019.

Rating: 3: Weak Reject
Experience Assessment: I have published in this field for several years.
Review Assessment: Thoroughness In Paper Reading: I read the paper thoroughly.
Review Assessment: Checking Correctness Of Derivations And Theory: I carefully checked the derivations and theory.
Review Assessment: Checking Correctness Of Experiments: I carefully checked the experiments.

Reviewer recommendation: 3 (Reject)

7/20 sentences are requests

This paper describes...	Structuring
...	Structuring
...	Structuring
...	Structuring
... was quite novel ...	Evaluative
...	Fact
...	Evaluative
... could highlight more recent work...	Request
...	Evaluative
I did appreciate...	Evaluative
...	Structuring
... benefit from more detail...	Request
...	Request
...	Fact
...	Structuring
...	Request
Additional citations suggested:	Structuring
[1]	Request
[2]	Request
[3]	Request

Fine-grained request labels

Evaluative

Request

Fact

Structuring

Social

Fine-grained request labels

Evaluative

Request

Fact

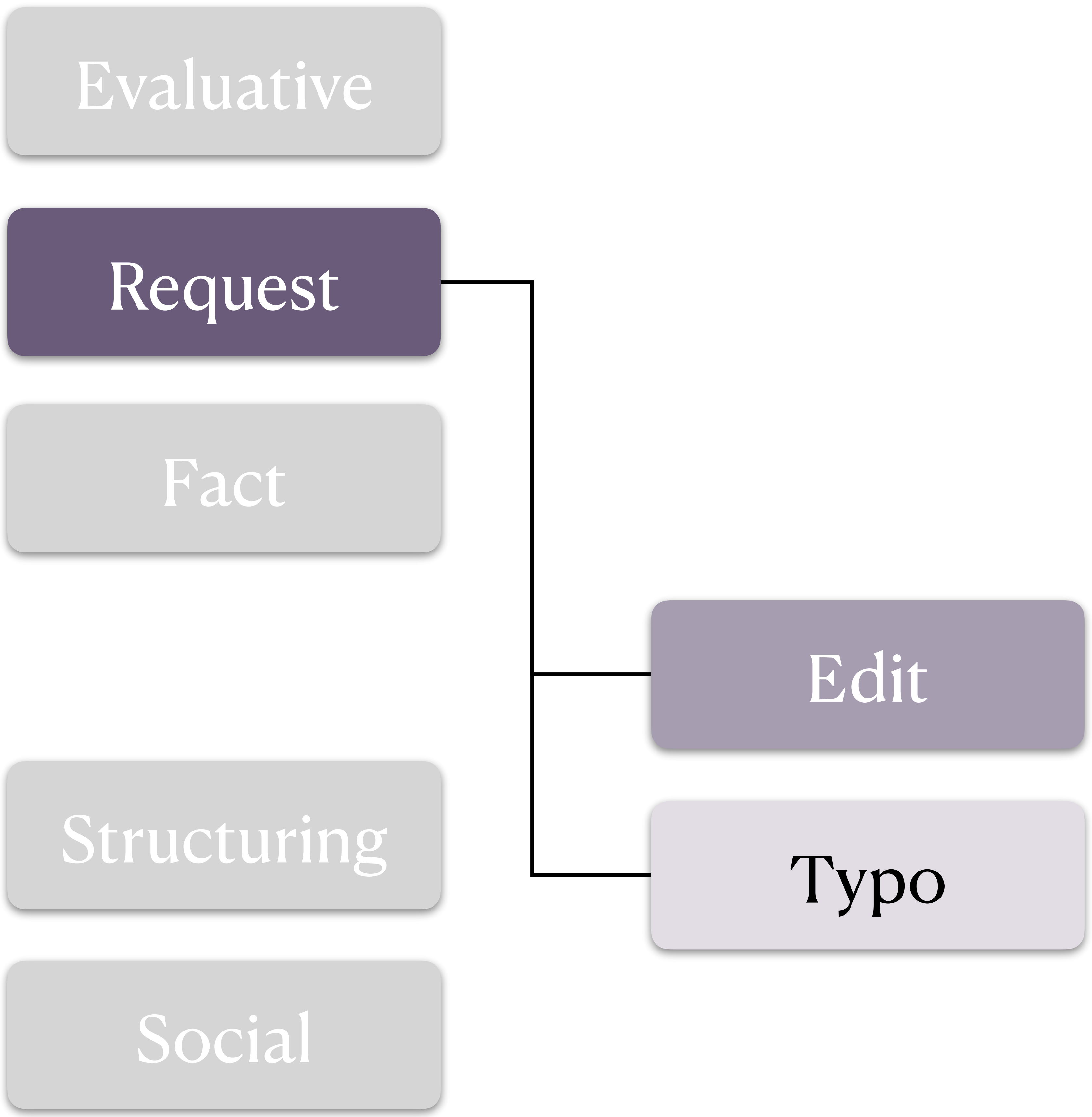
Structuring

Social

Typo

Capitalize: “section” -> “Section”

Fine-grained request labels



This restriction should be noted in the paper.

Capitalize: “section” -> “Section”

Fine-grained request labels

- Evaluative
- Request
  - Clarification — *Does the fMRI data exhibit time dependence?*
  - Edit — *This restriction should be noted in the paper.*
  - Typo — *Capitalize: “section” -> “Section”*
- Fact
- Structuring
- Social

Fine-grained request labels

- Evaluative
- Request
  - Explanation — *Won't we eventually expect exploding gradients…*
  - Clarification — *Does the fMRI data exhibit time dependence?*
  - Edit — *This restriction should be noted in the paper.*
  - Typo — *Capitalize: "section" -> "Section"*
- Fact
- Structuring
- Social

Fine-grained request labels

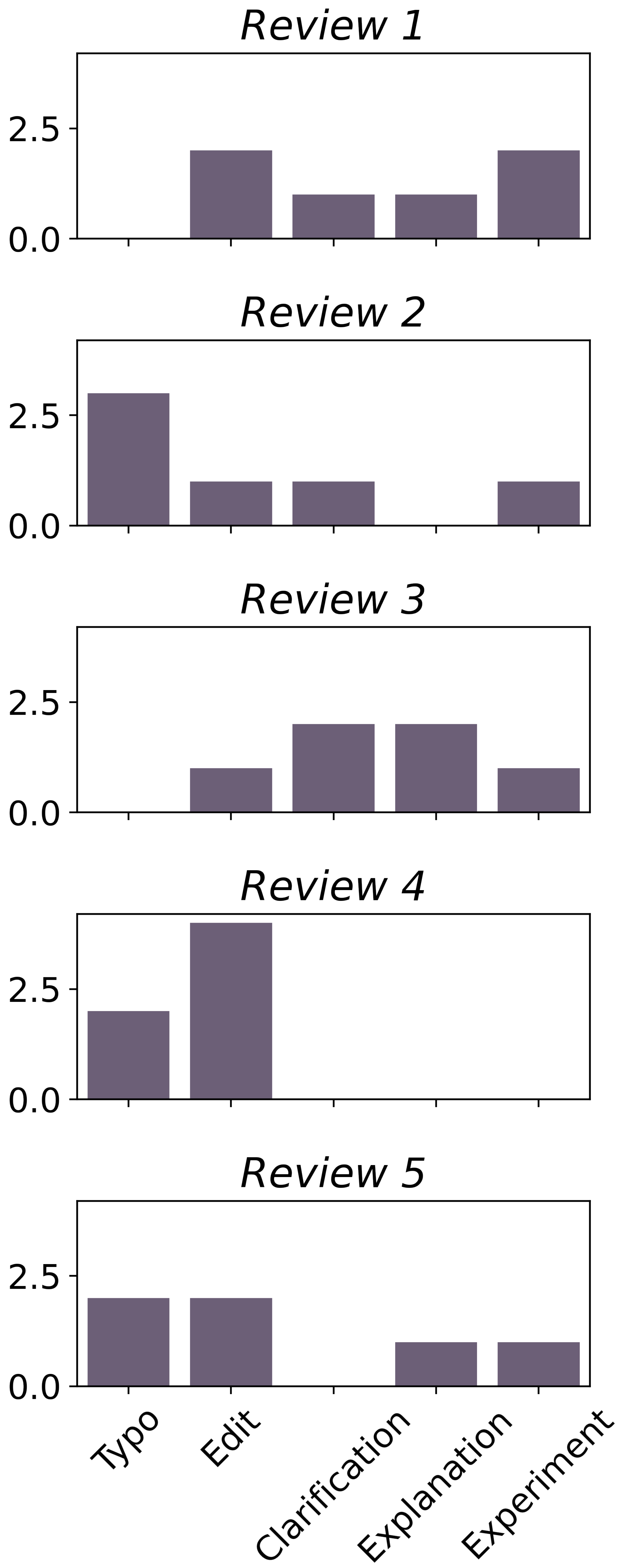
- Evaluative
- **Request**
  - Experiment — *It may be worth adding a comparison to that work…*
  - Explanation — *Won't we eventually expect exploding gradients…*
  - Clarification — *Does the fMRI data exhibit time dependence?*
  - Edit — *This restriction should be noted in the paper.*
  - Typo — *Capitalize: "section" -> "Section"*
- Fact
- Structuring
- Social

Reviewer recommendation: 3 (Reject)

6 request statements per review

Reviewer recommendation: 3 (Reject)

6 request statements per review



What about:

What about:

Questions based on a misunderstanding?

What about:

Questions based on a misunderstanding?

Requests for experiments that are just out of scope?

What about:

Questions based on a misunderstanding?

Requests for experiments that are just out of scope?

Idea: Use the rebuttal, which contains authors' opinion on the review

Sentence-level rebuttal labels

Sentence-level rebuttal labels

Alignment: context sentences in the review

Sentence-level rebuttal labels

Alignment: context sentences in the review

16 response types

Sentence-level rebuttal labels

Alignment: context sentences in the review

16 response types

Two categories: *concur* and *dispute*

Rebuttal stance: criticizing the criticism

Rebuttal stance: criticizing the criticism

Request

Add comparison to the following baselines...

Rebuttal stance: criticizing the criticism

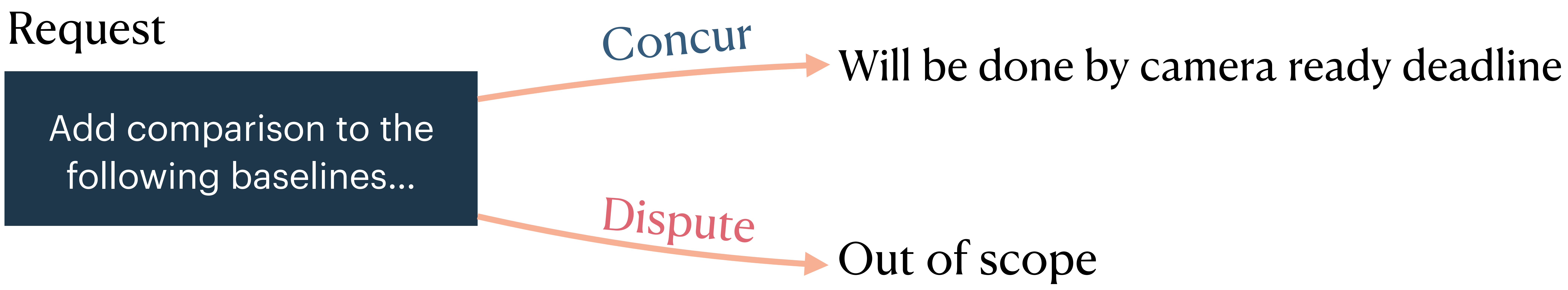
Request

Add comparison to the following baselines...

Concur

Will be done by camera ready deadline

Rebuttal stance: criticizing the criticism



Rebuttal stance: criticizing the criticism

Request

Add comparison to the following baselines...

Concur → Will be done by camera ready deadline

Dispute → Out of scope

Evaluative

This method might not be practically useful, since...

Rebuttal stance: criticizing the criticism

Request

Add comparison to the following baselines...

Concur → Will be done by camera ready deadline

Dispute → Out of scope

Evaluative

This method might not be practically useful, since...

Concur → Concede criticism

Rebuttal stance: criticizing the criticism

Request

Add comparison to the following baselines...

- Concur → Will be done by camera ready deadline
- Dispute → Out of scope

Evaluative

This method might not be practically useful, since...

- Concur → Concede criticism
- Dispute → Mitigate criticism

Agreeability

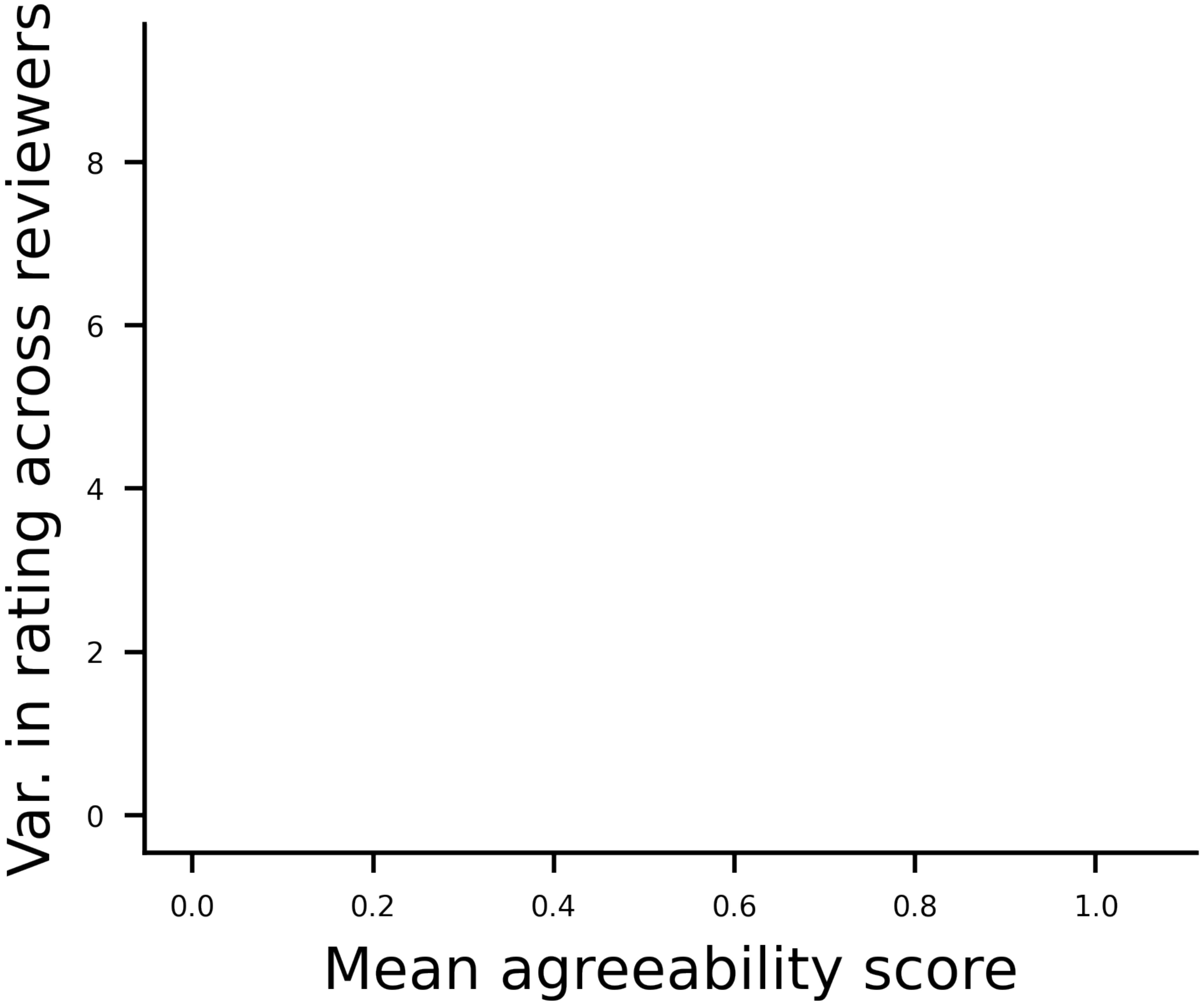
Out of argumentative statements, % of time author accepts premise

$$agreeability = \frac{n_{concur}}{n_{concur} + n_{dispute}}$$

Agreeability

Agreeability reveals controversies not apparent from score variance.

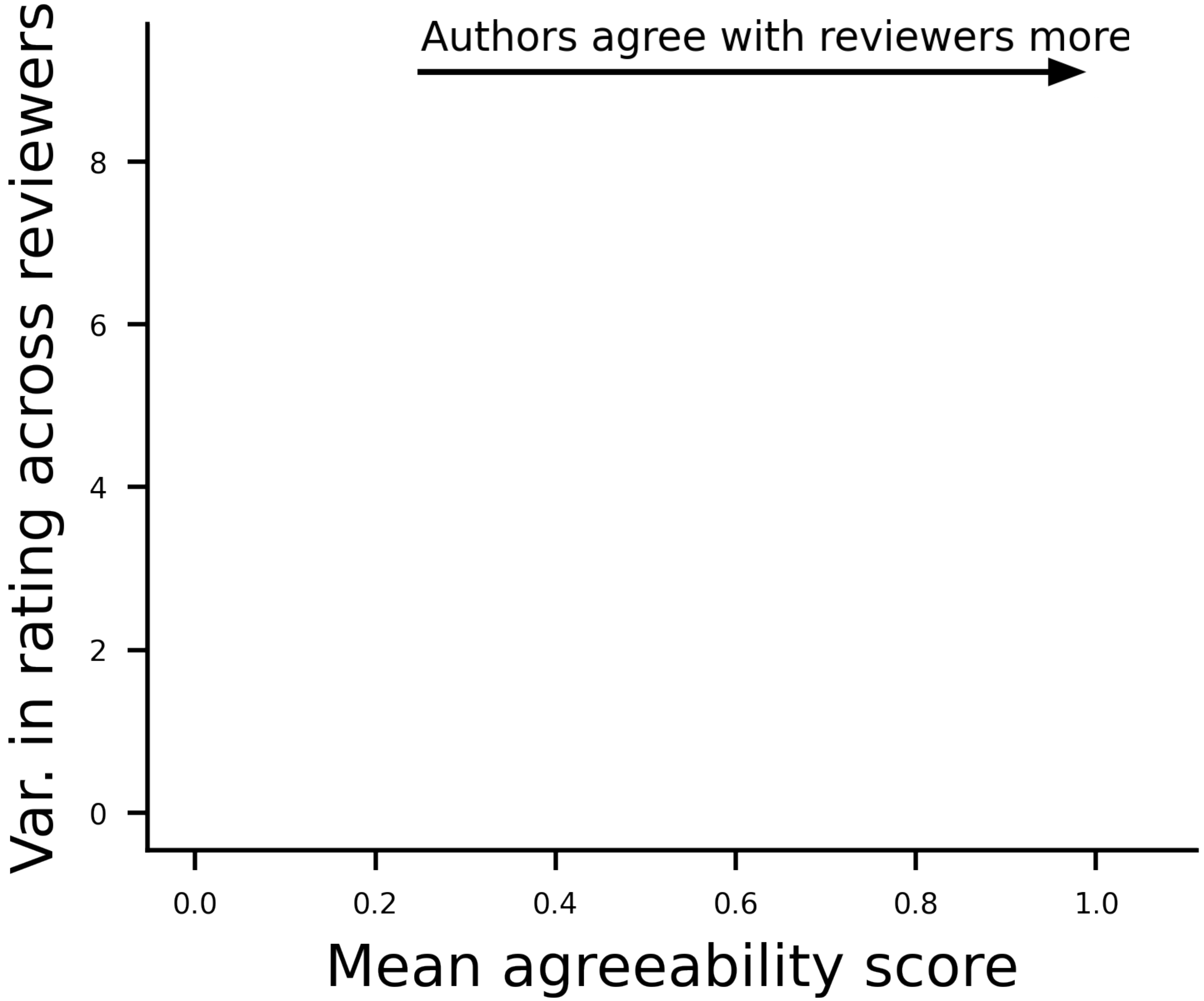
$$agreeability = \frac{n_{concur}}{n_{concur} + n_{dispute}}$$



Agreeability

Agreeability reveals controversies not apparent from score variance.

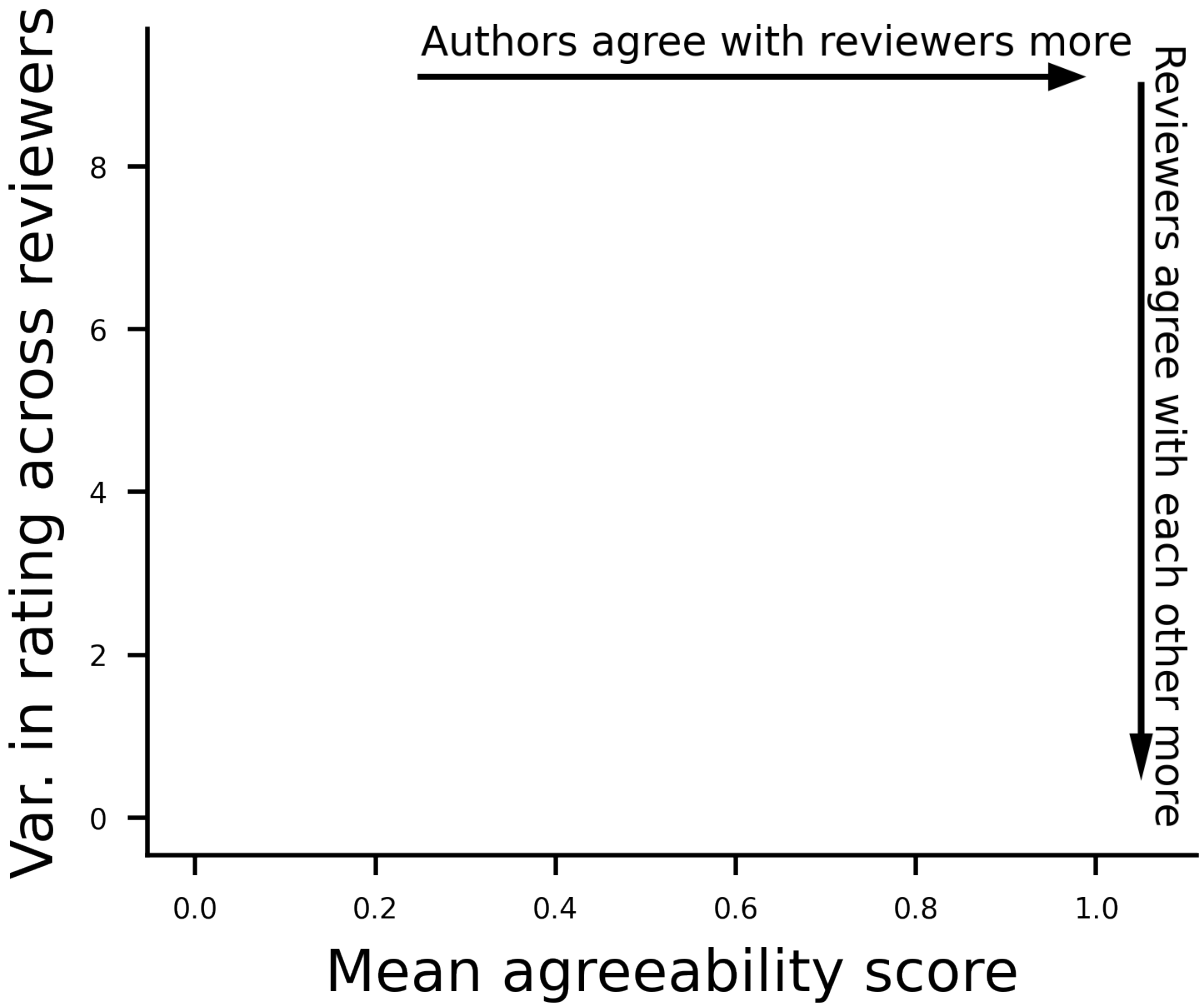
$$agreeability = \frac{n_{concur}}{n_{concur} + n_{dispute}}$$



Agreeability

Agreeability reveals controversies not apparent from score variance.

$$agreeability = \frac{n_{concur}}{n_{concur} + n_{dispute}}$$



Agreeability

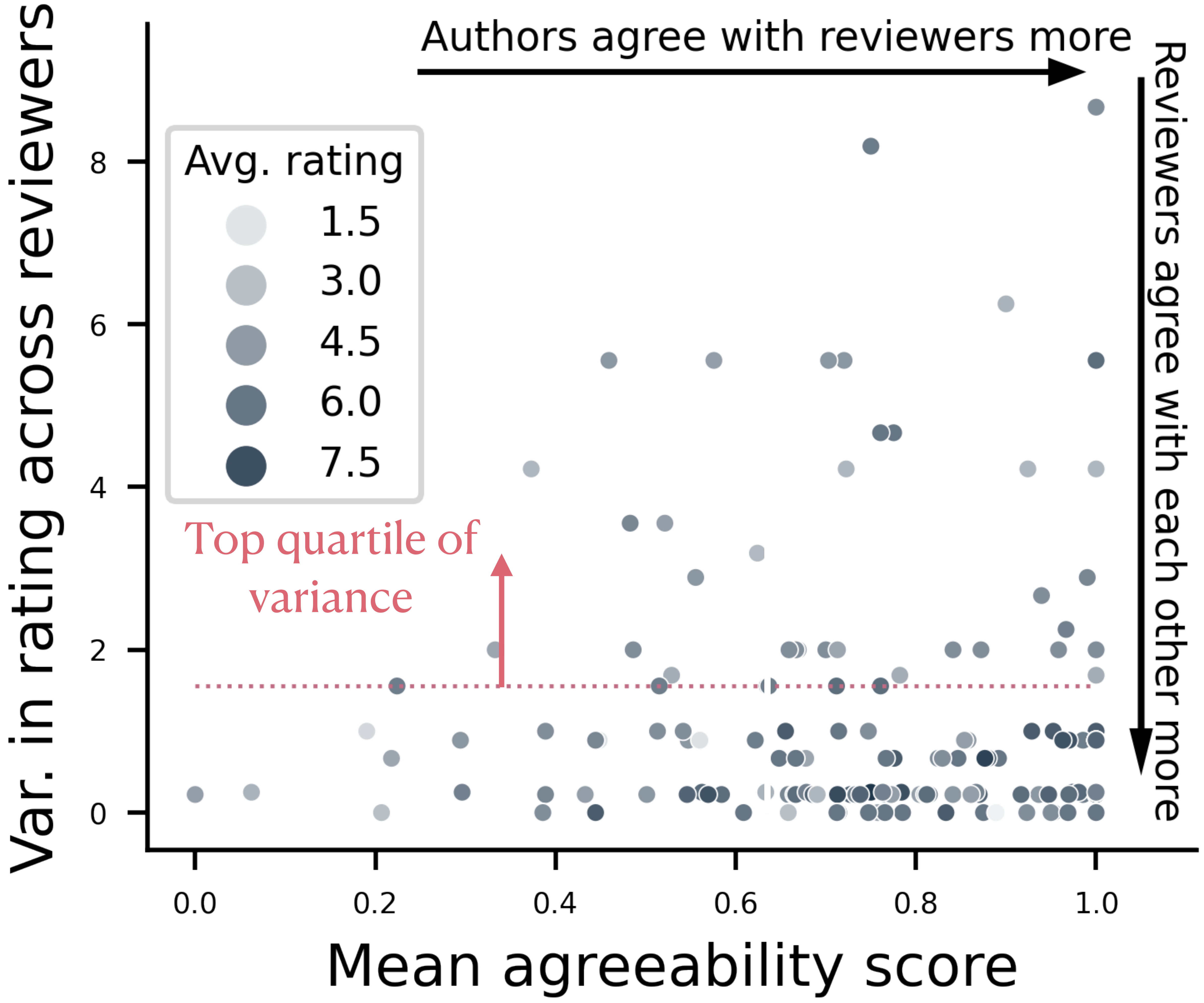
Agreeability reveals controversies not apparent from score variance.

$$agreeability = \frac{n_{concur}}{n_{concur} + n_{dispute}}$$

Agreeability

Agreeability reveals controversies not apparent from score variance.

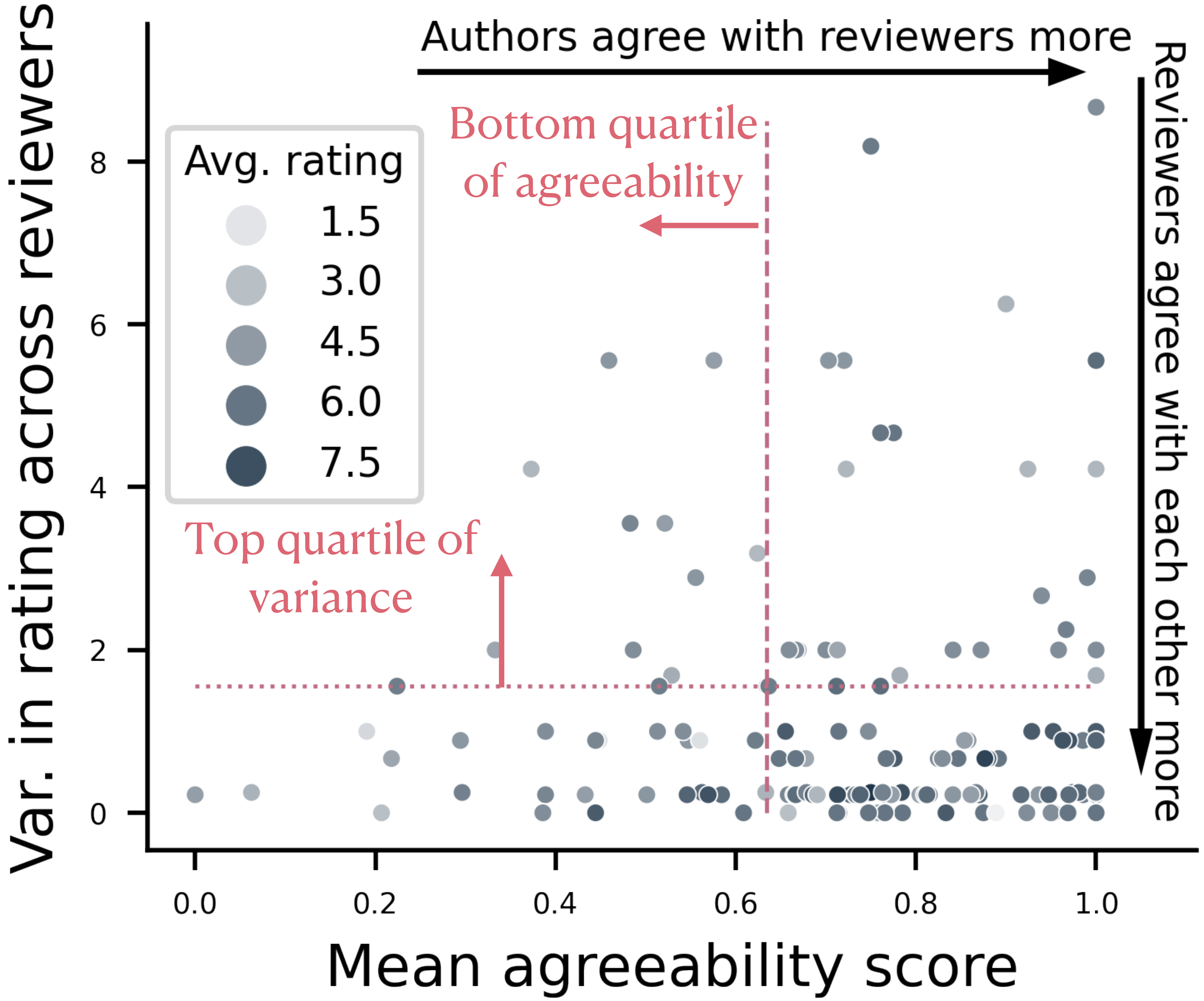
$$agreeability = \frac{n_{concur}}{n_{concur} + n_{dispute}}$$



Agreeability

Agreeability reveals controversies not apparent from score variance.

$$agreeability = \frac{n_{concur}}{n_{concur} + n_{dispute}}$$

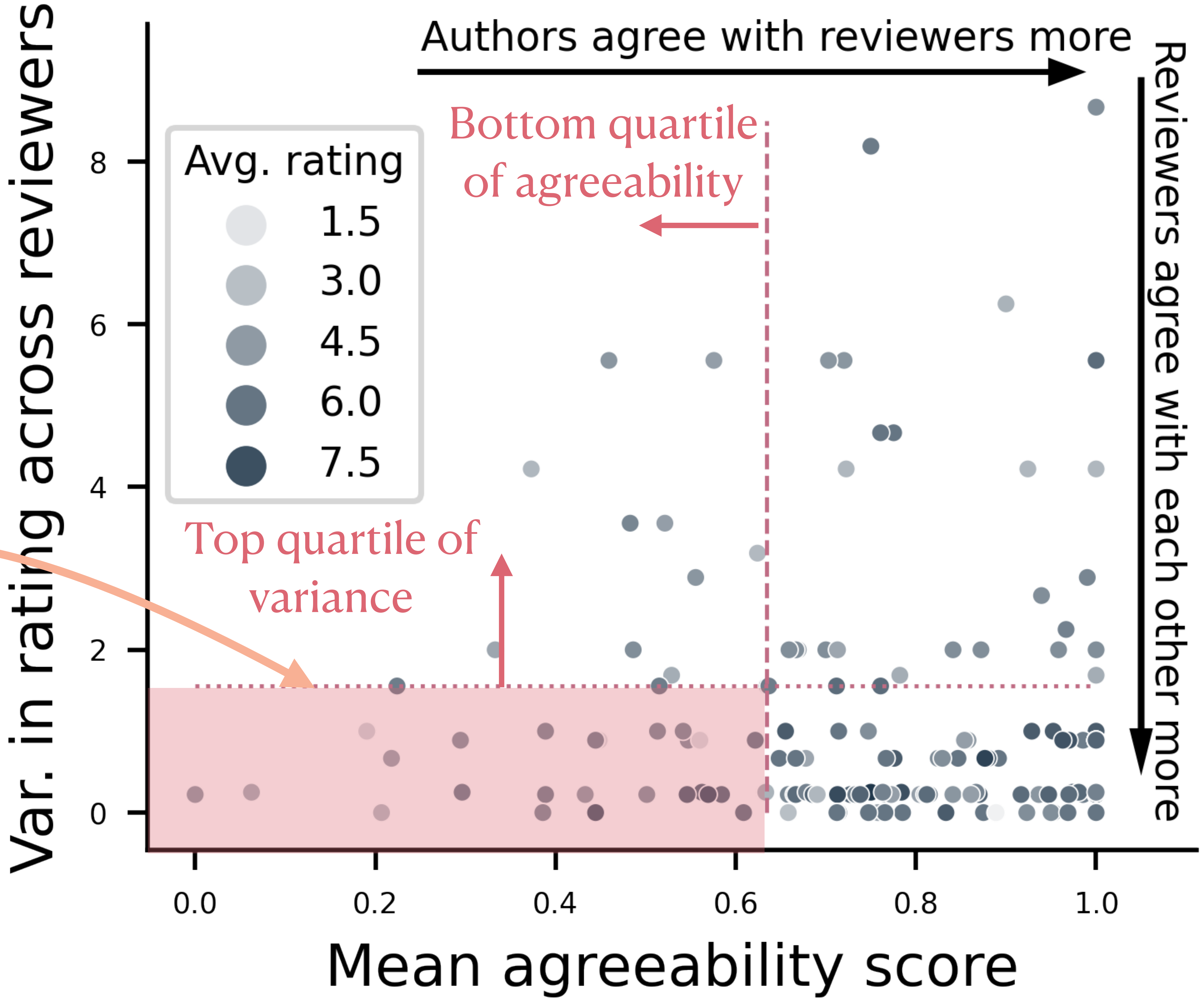


Agreeability

Agreeability reveals controversies not apparent from score variance.

$$agreeability = \frac{n_{concur}}{n_{concur} + n_{dispute}}$$

28/159 papers



DISAPERE

A dataset revealing nuances of peer review discussions

DISAPERE

A dataset revealing nuances of peer review discussions

How do we use these labels?

DISAPERE

A dataset revealing nuances of peer review discussions

How do we use these labels?

- Post-hoc analysis
- Designing policies and interfaces

DISAPERE

A dataset revealing nuances of peer review discussions

How do we use these labels?

- Post-hoc analysis
- Designing policies and interfaces

What about other (non-ML) domains?

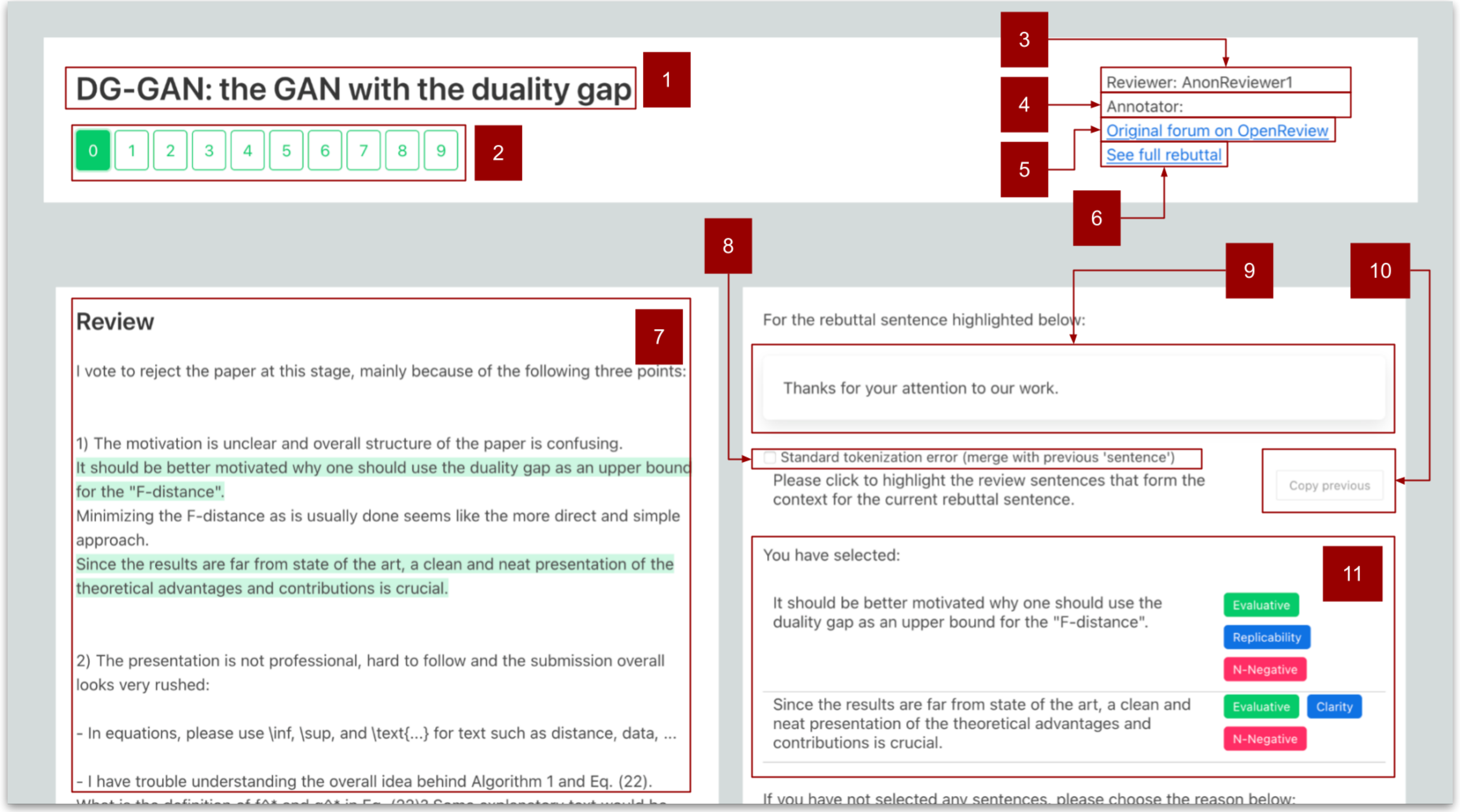
DISAPERE

A dataset revealing nuances of peer review discussions

How do we use these labels?

Post-hoc analysis

Designing policies and interfaces

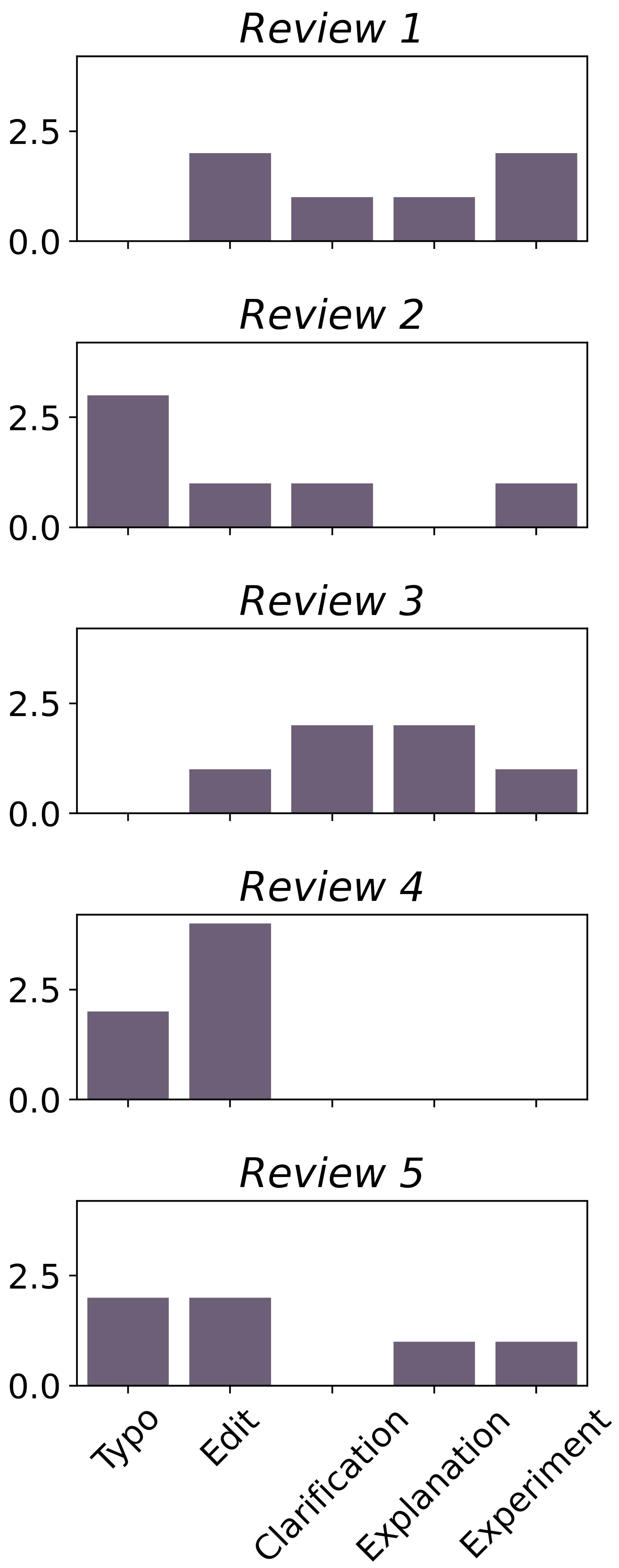


What about other (non-ML) domains?

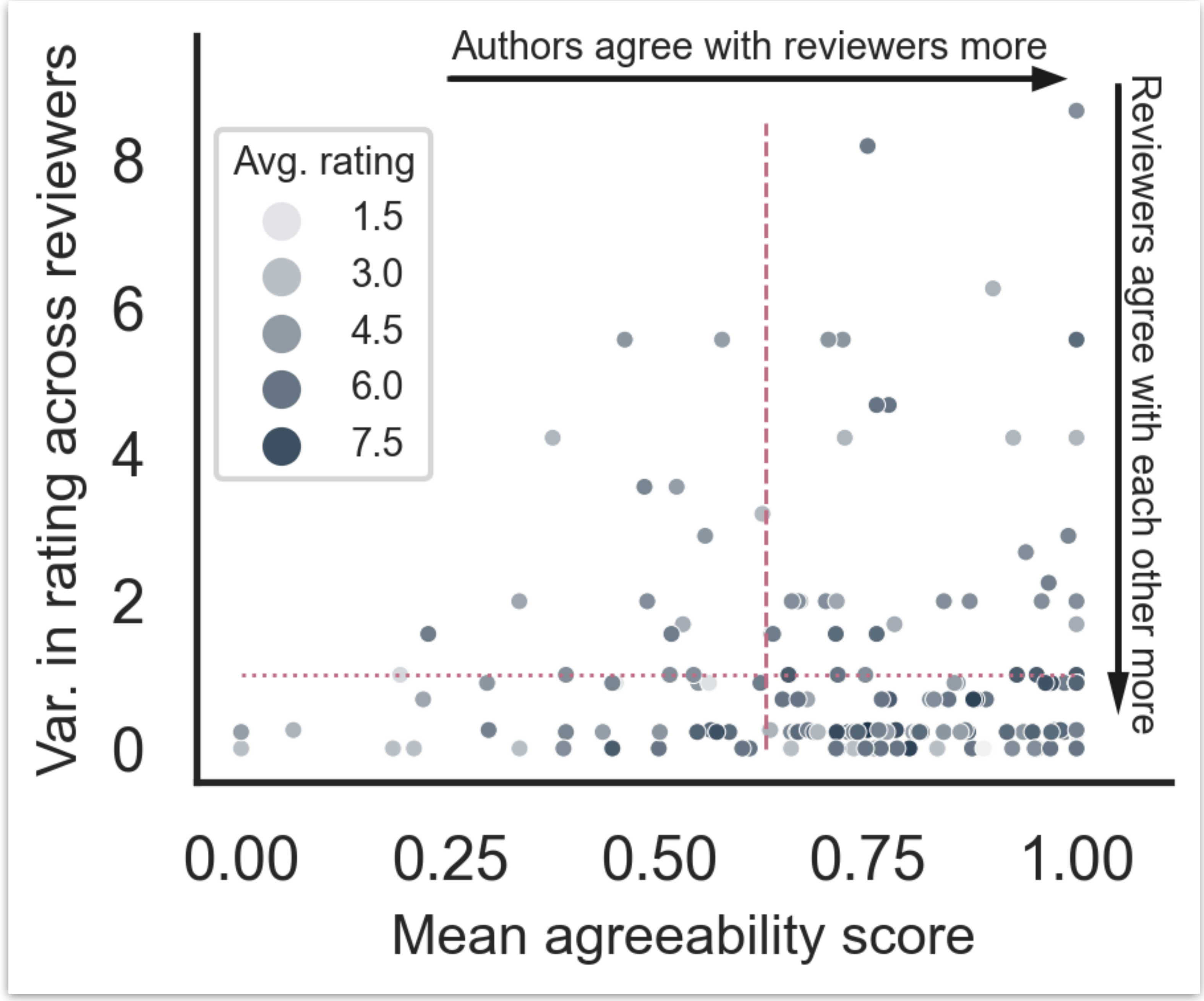
www.github.com/nnkennard/DISAPERE

Thank you!

Tim O'Gorman, Rajarshi Das, Akshay Sharma , Chhandak Bagchi
Matthew Clinton , Pranay Kumar Yelugam, Hamed Zamani, Andrew McCallum

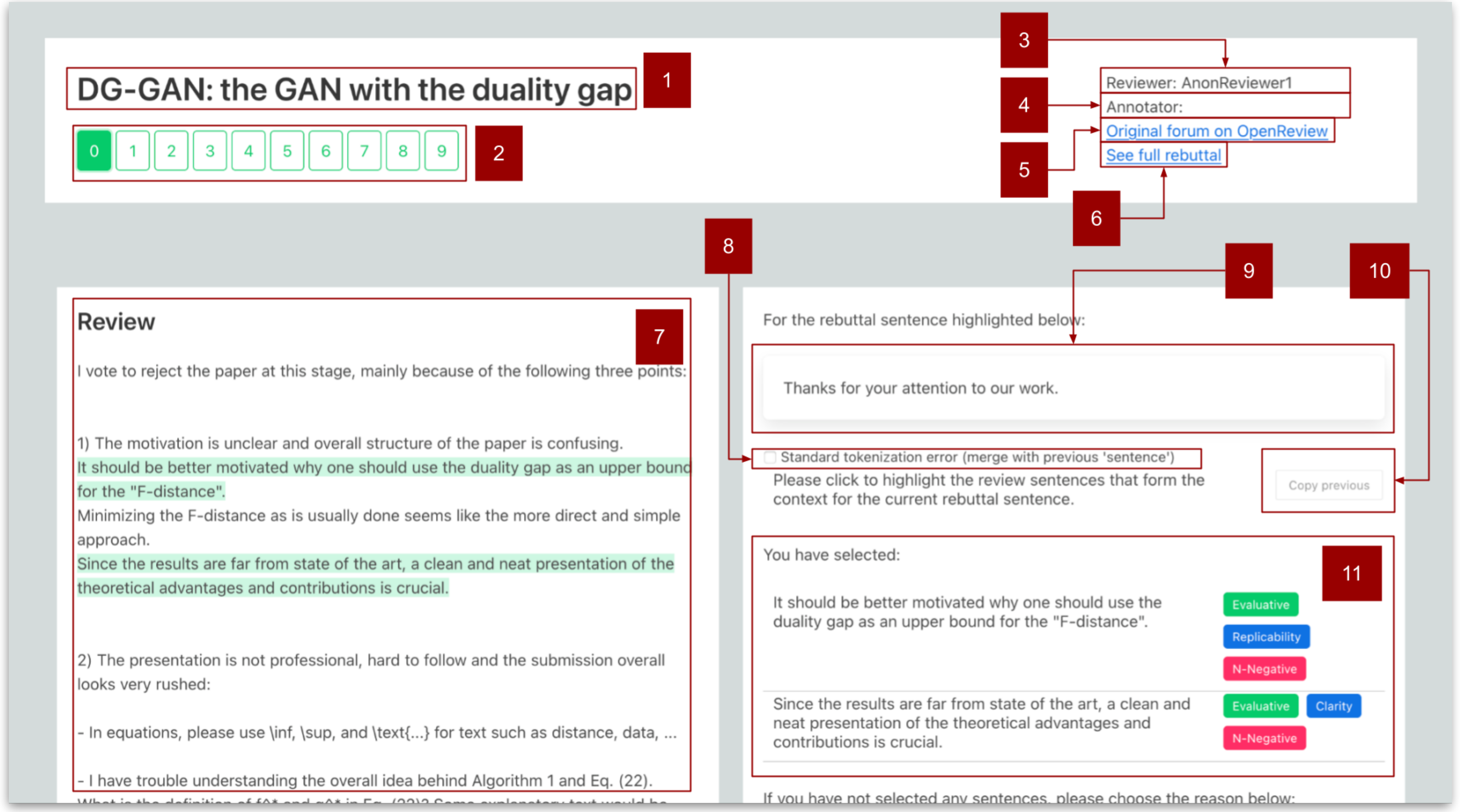


Was the feedback **constructive**?



Agreeability

Was the feedback **applicable**?



Software

www.github.com/nnkennard/DISAPERE

kennard@cs.umass.edu

@nnkennard