Fallacious Argument Template Instantiation Guidelines V2.0

Anonymous Author

1 Overview

A *fallacy* is an invalid or weak argument supported by unsound reasoning. The presence of fallacies has emerged as a significant concern due to their proneness to spread misleading and manipulative information [1]. The frequent occurrence of fallacies is due to a lack of proficiency in critical thinking skills. In situations where individuals fail to engage in systematic and analytical thinking, logical errors and inconsistencies often emerge, giving rise to fallacious reasoning patterns. The absence of critical thinking courses inside the school curriculum contributes to a lack of ability for students to construct an argument devoid of fallacies. Without structured instruction in critical thinking, students may struggle to notice logical flaws and develop the skills necessary for sound argumentation [2]. For instance, let us consider the following example:

- I know five people from Kentucky. They are all racists. Therefore, Kentuckians are racist.

This example contains a fallacy because it claims that Kentuckians are racist based on five racist people from Kentucky. It commits a *faulty generalization* (see Section 3.2 for more details) fallacy since it generalizes all Kentuckians based on small samples.

Traditional fallacy recognition systems focus primarily on identifying the fallacy type given a sentence [3, 4]; however, they do not go deeper into exploring and understanding why the fallacy has been committed. Suppose the previous example is identified as a faulty generalization and explains that "5 people from Kentucky generalized to conclude all Kentuckians are racist" this helps clarify the nature of the fallacy. Knowing the underlying reasons why a fallacy is committed can help students understand how to avoid committing fallacies in the future [2].

Towards helping students understand the components within their arguments that create a fallacy, this annotation task will consist of annotating a given fallacious argument. Using an inventory of fallacious argument templates, the first step is to choose the right argument template based on the fallacy types and then fill an appropriate event or entity into the slot-fillers of the chosen template. The results of this annotation will contribute towards explaining fallacies in finer detail.

2 Fallacy Template Components

In this Section, we discuss several components necessary for instantiating fallacious templates. We begin by discussing a popular argumentation scheme used as motivation for creating our patterns followed by the specific components.

2.1 Argument from Consequences

We adopted Argument from Consequences based on the argumentation scheme by [5] to develop the template, as Argument from Consequence is a widely used argumentation scheme [6]. In general, this scheme consists of 2 types: arguments from positive consequences and arguments from negative consequences.

Argument From Positive Consequences

- *Premise*: If A is brought about, good consequences will plausibly occur.
- *Conclusion*: Therefore, A should be brought about.

Argument From Negative Consequences

- *Premise*: If A is brought about, then bad consequences will occur.
- *Conclusion*: Therefore, A should not be brought about.

To evaluate this argument scheme, 3 critical questions need careful consideration.

Critical Questions

- **CQ1:** How strong is the likelihood that the cited consequences will (may, must) occur?
- **CQ2:** What evidence supports the claim that the cited consequences will (may, must) occur, and is it sufficient to support the strength of the claim adequately?
- **CQ3:** Are there other opposite consequences (bad as opposed to good, for example) that should be taken into account?

Suppose any of the critical questions cannot be answered, the argument is considered weak and potentially contains a fallacy [7]. Due to focusing on describing fallacy, we only adopt critical question 1 (CQ1) and critical question 2 (CQ2) in our templates.

2.2 Argumentative Components

Claim is a proportion that is used as a conclusion of an argument.

Premise P is a proportion that is used to support or attack the conclusion.

3 Fallacy Template Inventory

In this section, we discuss the inventory of fallacy templates we create which are each inspired by Argument from Consequence.

3.1 General Properties

3.1.1 Relations

Promote refers to activation of something. For example:

- "study makes you smart."

This example illustrates that studying **promote** smartness or that studying will activate smartness.

Suppress refers to the inactivation. For example:

- "studying takes away free time ."

This example illustrates that studying suppresses free time or free time is inactivated by studying.

3.1.2 Sentiment

Sentiment refers to the feeling an individual has towards something. In the case of our guidelines, we focus on the sentiment an author has towards both entities and events. More specifically, there are 2 types of sentiment we focus on:

Good sentiment is associated with positive attributes, as demonstrated in the **promote** example, where study and smart are both thought of as **good** by the author.

Bad sentiment is associated with negative attributes, as demonstrated in the **suppress** example, where study is considered as **bad** because the author believes that it **suppresses** free time which has a **good** sentiment.

3.1.3 Template Component

We decide to extend the premise based on the critical question to describe the fallacy.

Premise P serves as an argumentative component premise where the proportion is explained through **promote** and **suppress** relation. Premise P is used for supporting **Conclusion**

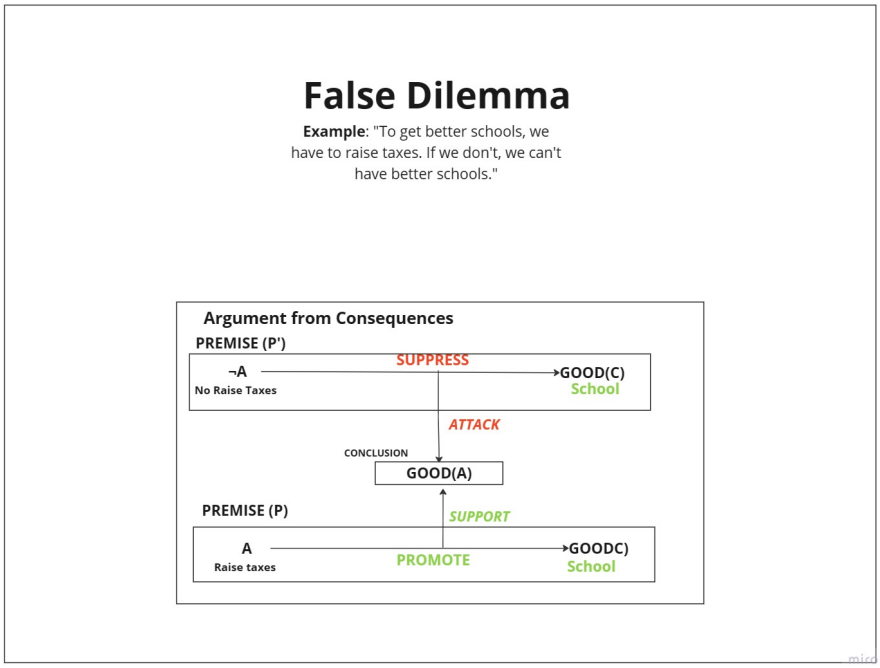
Premise P' serves as an argumentative component premise that is used to support premise *P*. However, for the *false dilemma*, this is used to support **Conclusion**.

Conclusion serves as an argumentative component claim supported by Premise *P*. However, for the *false dilemma*, **conclusion** is also supported by Premise *P'*.

3.1.4 Slot-Fillers

Each fallacy template is accompanied with slot-fillers, which consist of either an event or entity. An example is shown below:

- "To get better schools, we have to raise taxes. If we don't, we can't have better schools."



Ideally, we would like to exhaustively annotate all slot-fillers and templates for a given sentence; however, this introduces complexity into the annotation, the slot-filler could change which template is chosen. To address this, we first determine if an entity can be used as a slot-filler while maintaining the Argument from Consequences structure; otherwise, an entity may be chosen.

In the above example, *A* would be *raise taxes*, as the entity *taxes* alone implies that they are not already brought about. As for *C*, two possible options are *schools* (i.e., entity) or *can't have better schools* (i.e., event). If choosing the first option, we would say that schools are promoted, whereas the second option

would be suppressed as a result of raising taxes, hence two separate templates would result. For this specific case, we would choose *C=schools*, as we prefer entities over events, thus in the above example, template 1 would be selected.

3.2 Faulty Generalization

3.2.1 Definition

A faulty generalization fallacy occurs when an argument applies a belief to a large population without having a large enough sample to do so. Other terms for faulty generalization include 'hasty generalization'

3.2.2 Example

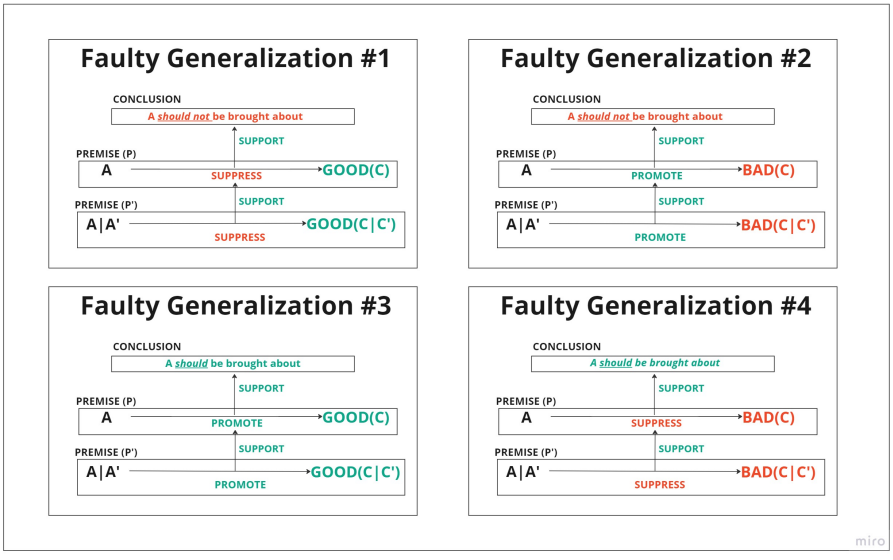
The following example demonstrates an argument which has committed a faulty generalization.

- "My friend swears the mechanic at that shop overcharged her last week and after looking at her invoice it seems she's right. So don't take your business there, as you'll definitely get ripped off"

This argument commits a faulty generalization because it generalizes that everyone will get overcharged if doing business with that shop because my friend got overcharged when doing business with that shop. Then, it also generalizes that the shop is bad only because a mechanic overcharged my friend.

3.2.3 Faulty Generalization Fallacy Templates

We create an inventory of 4 fallacy templates for representing faulty generalization.



Each template concludes with either a negative consequence (i.e., templates 1 and 2) or a positive consequence (i.e., templates 3 and 4). Each template consists of the following elements:

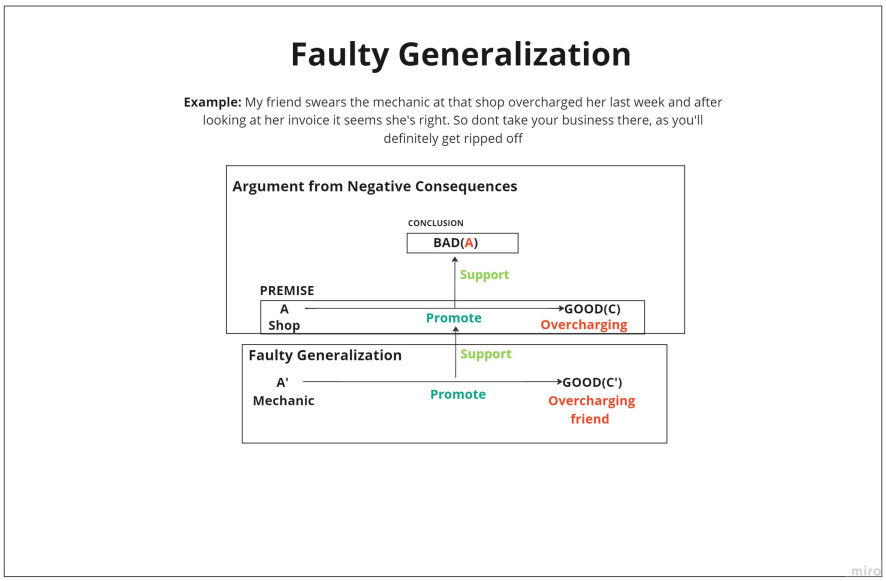
- **A**: A slot-filler inside premisel P and can be inside premise P' (from the claim) that will promote/suppress a slot-filler C (from the premise) and this is equivalent to the conclusion (A=**bad** is equivalent to A *should not be brought about*)
- **C**: A slot-filler inside premise P and can be inside premise P' (from the premise) which is a consequence of a slot-filler A (from the claim)
- **A'**: A slot-filler inside premise P' (from the premise) that will promote/suppress a slot-filler C/C' (from the premise) where A' is a subset of A
- **C'**: A slot-filler inside premise P' (from the premise) which is a consequence of a slot-fillers A (from the claim) or A' (from the premise) where C' is a subset of C

The premise P' provides a slot-filler selection such as A or A' and C or C' which align with the provided argument. Nevertheless, due to a faulty generalization, there is an occurrence where at least one slot-filler between A' or C' is chosen.

3.2.4 Template Instantiation Example

To exemplify template instantiation, we utilize the example from above:

- "My friend swears the mechanic at that shop overcharged her last week and after looking at her invoice it seems she's right. So don't take your business there, as you'll definitely get ripped off"



In this example, we can instantiate the main premise P with "A= Shop" and "C= Overcharging" because the provided illustration claims that the shop overcharges people who do business there. Thus, a bad entity *shop* promotes a bad event *overcharging*. Therefore, *shop should not be brought about*.

Upon further examination, a faulty generalization is committed because the evidence "My friend swears the mechanic at that shop overcharged her last week" generalize *overcharging friend* into *overcharging* anyone who engages the business with that shop. It also concludes that the *shop* has a bad sentiment, solely based on the mechanic action that ripped off his friend. This analysis reveals that *overcharging my friend* is a subset of the broader concept of *overcharging*, while the *mechanic* action is a subset of the activities conducted by the *shop*.

This is why in premise P', both A' and C' can be selected where "A'= Mechanic" and "C'= Overcharging friend". Now the premise P' explains that bad entity A' promote bad event C'.

Therefore, a bad entity *mechanic* promotes a bad event *overcharging friend*, which supports the premise P that the *shop* promotes *overcharging*. This example is suitable for the template 2 structure. As a result, this template instantiation implies that 2 faulty generalizations occurred in this argument: A' generalizing A and C' generalizing C.

3.3 False Dilemma

3.3.1 Definition

A false dilemma fallacy is when incorrect limitations are made on the possible options in a scenario when there could be other options.

3.3.2 Example

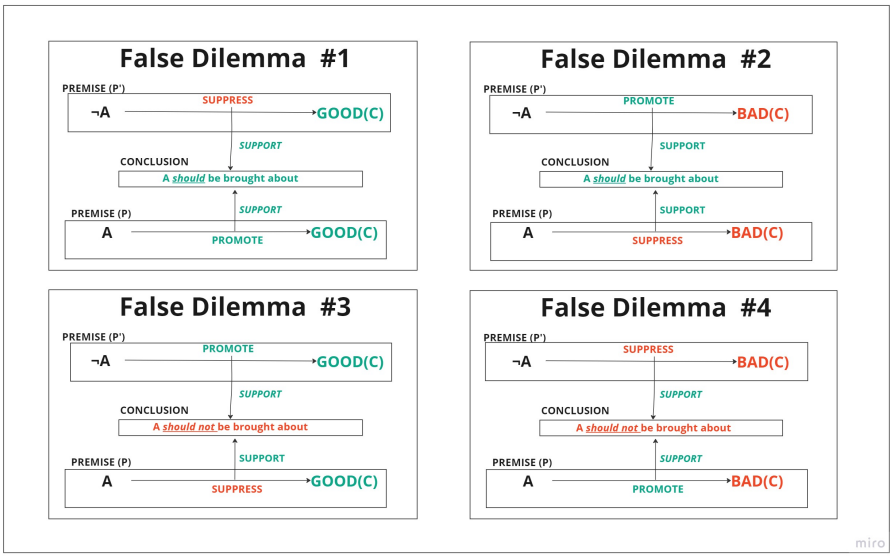
The following example demonstrates an argument which has committed a false dilemma.

- "We either have to cut taxes or leave a huge debt for our children."

This argument falls into a false dilemma fallacy because there is a restriction on the available choices to either *cut taxes* or *no cut taxes*, without considering any potential options. The absence of an explanation regarding the limitation of the options signifies that the argument is weak.

3.3.3 False Dilemma Templates

We create an inventory of 4 fallacy templates for representing false dilemma.



Each template concludes with either a positive consequence (i.e., templates 1 and 2) or a negative consequence (i.e., templates 3 and 4). Each template comprises of the following elements:

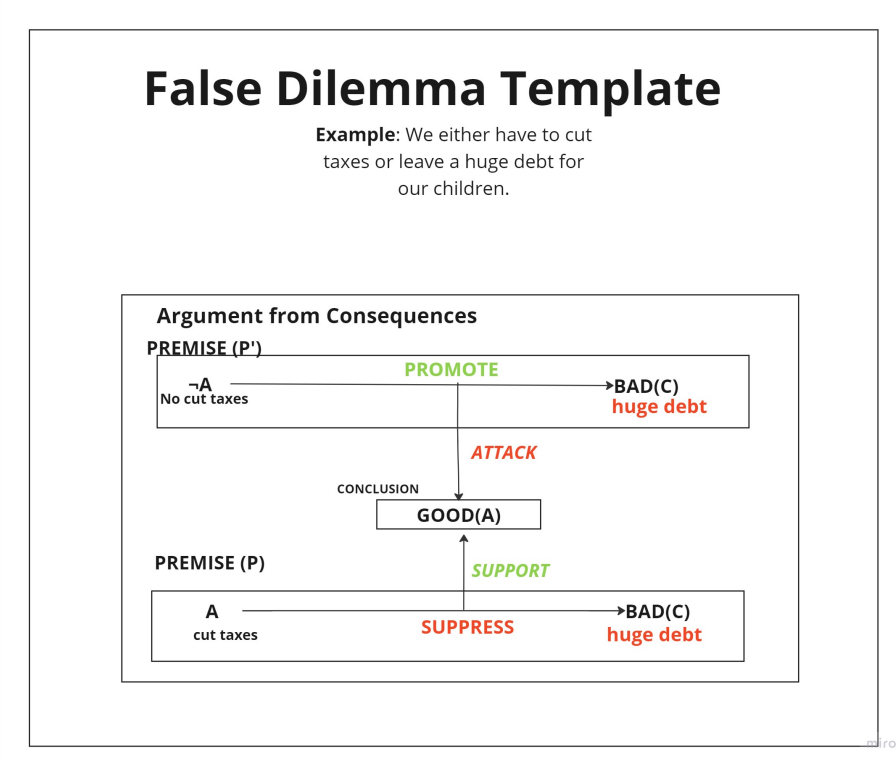
- **A**: A slot-filler inside premise P (from the claim) that will promote/suppress a slot-filler C (from the premise) and this is equivalent to the conclusion (A=**bad** is equivalent to A *should not be brought about*)
- **C**: A slot-filler inside premises P and P' (from the premise) which is a consequence of a slot-filler A (from the claim)
- **¬A**: A slot-filler inside premise P' of not A (i.e., A: Donate, ¬A: Not donate)

Within the template, there is a difference in sentiment between A and $\neg A$. Furthermore, the relations in premise P' consistently mirror the premise P due to the reverse type within the slot-fillers. For instance, if slot-filler A has a **bad** sentiment, **suppress** a **good** sentiment of slot-filler C, then $\neg A$ denotes a slot-filler that has a **good** sentiment and **promote** slot-filler C and vice versa. However, during annotation, the slot-fillers are restricted to A and C. Therefore, suppose A is already determined, $\neg A$ automatically represents the negation of A within the template.

3.3.4 Template Instantiation Example

To exemplify template instantiation, we utilize the example from above:

- "We either have to cut taxes or leave a huge debt for our children."



In this example, we can instantiate the premise P with "A= cut taxes" and "C= huge debt", consequently inside premise P', automatically instantiate "$\neg A$= No cut taxes" due to the claims that "either have to cut taxes or leave a huge debt". This implies that *cut taxes* is suppressing *huge debt*, while *no cut taxes* is promoting a *huge debt*. Therefore, *cut taxes should be brought about*. This example is suitable for the template 2 structure. As a result, this template instantiation implies that the following example is a false dilemma, as it restricts the available options to premises P and P'.

3.4 False Causality

3.4.1 Definition

A false causality fallacy occurs when an argument assumes that since two events are correlated, they must also have a cause and effect relationship.

3.4.2 Example

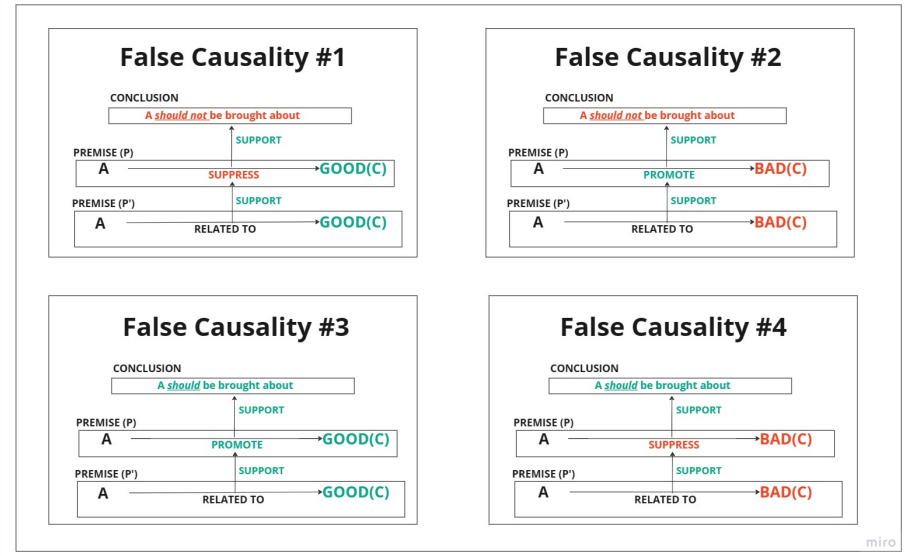
The following example demonstrates an argument which has committed a false causality.

- "The temperature has dropped this morning, and I also have a headache. The cold weather must be causing my headache."

This argument commits false causality because the argument illustrates the correlation of cause and effect between cold weather and headaches.

3.4.3 False Causality Template

We create an inventory of 4 fallacy templates for representing false causality.



Each template concludes with either a negative consequence (i.e., templates 1 and 2) or a positive consequence (i.e., templates 3 and 4). Each template comprises of the following elements:

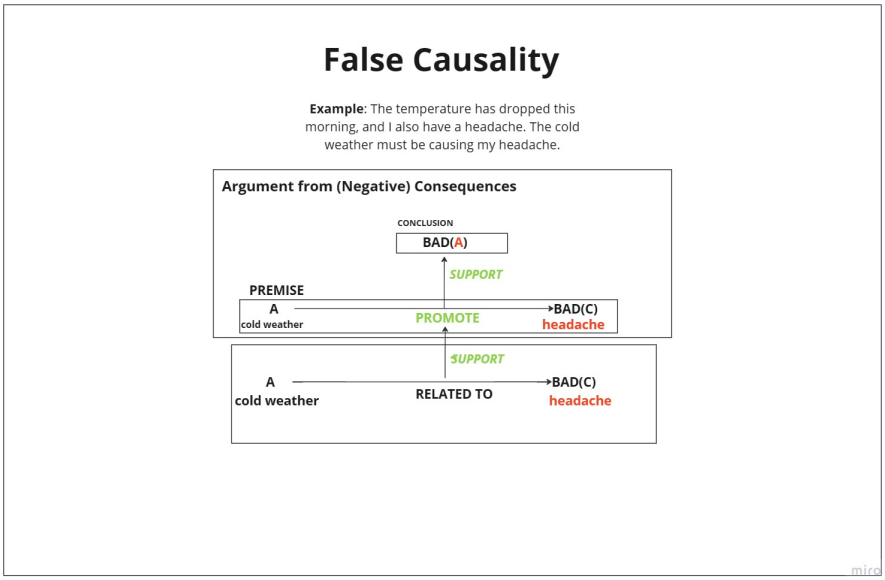
- **A**: A slot-filler inside premises *P* and *P'* (from the claim) that will promote/suppress a slot-filler C (from the premise) and this is equivalent to the conclusion (*A*=**bad** is equivalent to *A should not be brought about*)
- **C**: A slot-filler inside premises *P* and *P'* (from the premise) which is a consequence of a slot-filler *A* (from the claim)

Premise *P'* illustrates the correlation between slot-fillers *A* and *C*. Suppose the correlation exists and supports the cause and effect relation within the premise, the argument is considered as false causality.

3.4.4 Template Instantiation Example

To exemplify template instantiation, we utilize the example from above:

- "The temperature has dropped this morning, and I also have a headache. The cold weather must be causing my headache."



In this example, we can instantiate the premise *P* with "*A*= cold weather" and "*C*= headache" due to the claims that "the cold weather must be causing my headache". This implies a causal relationship, suggesting that *cold weather* promotes *headache*, resulting in a negative sentiment toward *cold weather*. Therefore, a bad event *cold weather* promotes a bad event *headache*. As a result, the conclusion is *cold weather should not be brought about*.

Investigating further into the premise *P'*, the relation between *cold weather* and *headache* serves as the basis of the cause and effect relationship. This example is suitable for the template 2 structure. As a result, this template instantiation implies that false causality is committed due to the relation between *A* and *C* that support the cause and effect relationship.

3.5 Fallacy of Credibility

3.5.1 Definition

A fallacy of credibility fallacy is when an appeal is made to some form of ethics, authority, or credibility.

3.5.2 Example

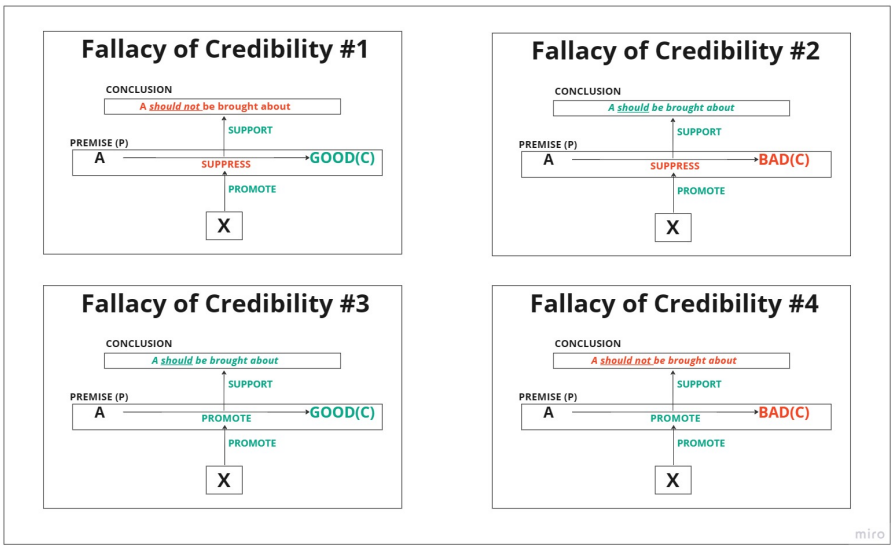
The following example demonstrates an argument which has committed a fallacy of credibility.

- "My uncle is a mechanic and he says you shouldn't spank children. He says it's ineffective."

This argument commits a fallacy of credibility by resorting to an appeal to authority, relying on advice from his uncle. However, the argument did not provide a further explanation regarding the ineffectiveness of child spanking.

3.5.3 Fallacy of Credibility Template

We create an inventory of 4 fallacy templates for representing fallacy of credibility.



Each template concludes with either a negative consequence (i.e., templates 1 and 4) or a positive consequence (i.e., templates 2 and 3). Each template consists of the following elements:

- **A**: A slot-filler inside premise P (from the claim) that will promote/suppress a slot-filler C (from the premise) and this is equivalent to the conclusion (A=**bad** is equivalent to *A should not be brought about*)

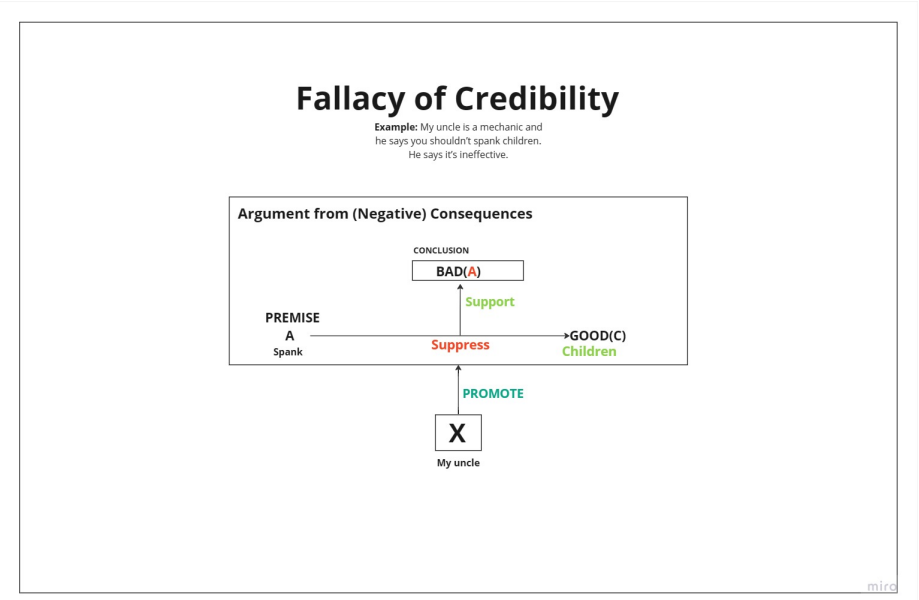
- **C**: A slot-filler inside premise P (from the premise) which is a consequence of a slot-filler A (from the claim)
- **X**: A person or group who appeals to the form of ethics, authority, or credibility.

If the conclusion should or should not be brought about based on a premise which is promoted by an appeal from X, the argument will be considered a fallacy of credibility.

3.5.4 Template Instantiation Example

To exemplify template instantiation, we utilize the example from above:

- "My uncle is a mechanic and he says you shouldn't spank children. He says it's ineffective."



In this example, we can instantiate the premise P with "A= Spank" and "C= Children" due to the claim that "you should not spank children". Thus, a bad event *spank* suppresses good entity *children*, leading to the conclusion that *spank should not be brought about*.

However, this example also reveals that the claim is spoken by his uncle who is a mechanic, also the uncle emphasizes that "it is ineffective". This implies the existence of an appeal to authority by the uncle. Therefore, we can instantiate the slot-filler X with "X= my uncle". Therefore, the premise was promoted by my uncle. This example is suitable for the template 1 structure. As a result, this template instantiation implies that the argument committed the fallacy of credibility due to the premise being promoted by X.

4 Implicit Elements

During annotation, we encourage annotators to please try to think of the sentence as if it were paraphrased as an Argument from Consequence argument. However, in doing so, annotators may encounter several implicit elements. We exemplify such instances below.

4.1 Argumentative Components

In certain instances, there may be an implicit conclusion or premise. Take the following example:

Let us consider the following example:

- "You should eat more green leafy vegetables, because Ms. Lord says so, and she is an expert on grammar and vocabulary."

This example committed a *fallacy of credibility* due to an appeal from Ms. Lord. However, this example has an implicit premise since the claim "you should eat more green leafy vegetables" does not provide any clear explanation regarding the consequence of eating more green leafy vegetables. For this case, try to be as explicit as possible. Hence, if we instantiate this example into the fallacy of credibility template, we select "$A=$ green leafy vegetables" promotes "$C=$ you" promoted by "$X=$ Ms. Lord" since *you* is the only explicit consequence even though it is vague.

4.2 Ingredients

One of the reasons that fallacy is hard to identify is due to the implicit ingredient inside the argument. The argument is deemed to possess an implicit ingredient if there is at least 1 argumentative ingredient that is not explicitly mentioned but can generally be inferred by the reader.

Suppose the sentiment associated with the slot-filler is subjective, leading to an uncertain or ambiguous relational connection. We consider that as Implicit Relation. For instance, let us consider the following example:

- "You can either support our police or Black Lives Matter."

In this *Fallacy of False Dilemma* example, the sentiment towards *police* and *Black Lives Matter* is subjective as it requires world knowledge. For this specific case, we encourage you to utilize the argument as much as possible and its explicit contents without bringing in background knowledge towards the topic. Therefore, since the author explicitly writes *support* in *support our police*, it indicates that *support our police* has good sentiment, whereas **Black Lives Matter** would have bad sentiment based on the notion that this is a *Fallacy of False Dilemma* argument.

5 Non-Argument from Consequences

As mentioned a priori, during annotation, please try to think of the sentence as if it were paraphrased as an Argument from Consequence argument. In various cases, there may be arguments that still cannot be rewritten as an Argument from Consequences and thus not covered by our template inventory. We consider this as no consequence. There are several reasons that some arguments cannot be covered by the template

5.1 "Not Promote" and "Not Suppress"

Currently, fallacy templates can be represented using only *promote* and *suppress* relations (see Section 3.1.1. However, a sentence relation could also be *not promote* or *not suppress*.

An example of such relation is shown below:

- "There were wonderful psychologists who passed away several decades ago. If they could be effective in what they did without reading any of the studies or other articles that have been published in the last several decades, there's no need for me to read any of those works in order to be effective."

The above example mentions that *reading* does *not promote effectiveness*. We note that "not promote" relation is not equivalent to *suppress* relation, and vice versa.

For such instances, we kindly ask that you do not annotate them using a fallacy template. If possible, please use the Note column to indicate this.

6 Procedure of the Annotation

Below are the following procedures for the annotation process:

- Carefully read the given sentence and the associated fallacy type.
- Identify the potential templates and the corresponding slot-fillers that need to be filled.
- Select the most suitable template and fill the slot-fillers.
- For the purpose of this annotation, please assume for the given sentence that the given fallacy type has been committed.

6.1 Key Considerations in Annotation

Please consider the following points during annotation:

- If necessary, even if an argument from consequence structure is not present in the argument, try to think about it as though it were paraphrased to fit the scheme. This is the case, especially for False Dilemma arguments.

- When filling the slot-filler, be as simple as possible.
- Occasionally, the use of an entity as a slot-filler can make the conclusion of the template seem incomprehensible (e.g., A=football players). To mitigate this issue, we encourage annotators to think about the conclusion as "*A should be promoted*" or "*A should be suppressed*".
- For *False Dilemma*, if there exists more than one slot-filler A, fill the slot-filler A for the event or entity that is shown first.
- When filling the slot-filler, we encourage using explicit events or entities as much as possible.
- Despite the occurrence of implicit components, the explicit slot-filler should be consistent with the original content of the argument (example in section 4.1)
- In the event one slot-filler is a subset of another (e.g., A= Allow to use my textbook in the medical exam, and C= medical exam), please try to determine if another slot-filler can be utilized; otherwise, please proceed with annotating with the subset as a slot-filler.
- When filling the slot-fillers, the entity is the top priority to be selected. In the event an entity cannot capture the essence of the original argument, please choose the event instead (Section 3.1.3).
- In *Fallacy of Credibility*, suppose multiple slot-fillers for X are available, select the slot-filler X that supports the main content of the argument.
- If you find that a consequence falls into the category of **Non-Argument from Consequence** (Section 5), please fill in the template as number 5 instead.
- If you are not 100% confident in your annotation, please specify how confident you are (%) along with your annotation.

References

[1] M. Hinton, *Evaluating the Language of Argument*, vol. 37. Springer Cham, 1 ed., 11 2020.

[2] B. Kim, *Critical Thinking*. Oklahoma State University Libraries, 2019.

[3] P. Goffredo, S. Haddadan, V. Vorakitphan, E. Cabrio, and S. Villata, "Fallacious argument classification in political debates," in *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI*, pp. 4143–4149, 2022.

[4] Z. Jin, A. Lalwani, T. Vaidhya, X. Shen, Y. Ding, Z. Lyu, M. Sachan, R. Mihalcea, and B. Schoelkopf, "Logical fallacy detection," in *Findings of the Association for Computational Linguistics: EMNLP 2022* (Y. Goldberg, Z. Kozareva, and Y. Zhang, eds.), (Abu Dhabi, United Arab Emirates), pp. 7180–7198, Association for Computational Linguistics, Dec. 2022.

[5] D. Walton, C. Reed, and F. Macagno, *Argumentation schemes*. Cambridge University Press, 2008.

[6] P. Reisert, N. Inoue, T. Kuribayashi, and K. Inui, "Feasible annotation scheme for capturing policy argument reasoning using argument templates," in *Proceedings of the 5th Workshop on Argument Mining*, pp. 79–89, 2018.

[7] D. Walton, *Informal logic: A pragmatic approach*. Cambridge University Press, 2008.