

## Additional Results for Experiment 2

In the following, we discuss the open-ended explanations provided by participants in response to justifying their rank of a statement in experiment 1. In our analysis, we found several explanation categories or groupings of similar explanations in terms of premises and conclusion. We refer to open-ended explanations by the term Other.

For statement 1, if Damian S is an aeroplane pilot then Damian S has visited different countries, the explanation category Other was included in the response of two participants. The first participant's explanation corresponds to category B&Other. This explanation means that the participant endorsed the premises, but not the conclusion and that they provided their own explanation. They provided the following explanation: "Not all pilots fly internationally". This explanation suggests a different form of reasoning to the classical logic interpretations of material implication statements. Moreover, their explanation suggests that there are exceptions to the class of typical pilots. Their reasoning is consistent with non-monotonic defeasible reasoning. For one, the participant could infer that some pilots might fly internationally and others might fly locally, within their country of origin. The second participant's explanation corresponds to the category Other. They provided the following explanation: "good". This response was vague, given that no other explanation was chosen. It is possible that this participant was hasty in completing the survey and chose the path of least resistance by selecting the shortest explanation - Other - and providing a neutral comment. Nonetheless, their explanation does not contradict any of the classical logic interpretations and thus does not warrant exclusion from our analysis of the statement's plausibility. For statement 2, if Rochelle P is a ballerina then Rochelle P does listen to classical music, the explanation category Other was included in the response of one participant. This participant's explanation corresponds to the category Other and provided the following information: "good". We note that this response was repeated by the same participant as for statement 1, and thus our previous analysis is valid for statement 2 as well. For statement 3, if Jacob B is a truck driver then Jacob B does drive at night, the explanation category Other was included in the response of two participants. The first participant's explanation corresponds to the category A&Other. This explanation means that the participant endorsed both the premises and the conclusion and that they provided their own explanation. They provided the following: "It's also possible that she's just a really bad lawyer and \*doesn't\* wear a suit but I'm going with chances are she wears one". This explanation suggests a different form of reason-

ing to the classical logic interpretations of material implication statements. Moreover, their explanation suggests that there are exceptions to the class of typical lawyers, e.g., “a really bad lawyer”. Their reasoning is consistent with non-monotonic reasoning. For one, this participant could infer that some lawyers work remotely and do not dress in a suit. Their explanation concludes with the remark, “but I’m going with chances are she wears one” which tells us that their overall belief is consistent with their choice of explanation A. The second participant’s explanation corresponds to the category Other. They provided the information: “good”. We note that this response was repeated by the same participant as statements 1 and 2, and thus our previous analysis is valid for statement 3 as well. For statement 6, if Zeeta M is a classical pianist then Zeeta M is a classical pianist, the explanation category Other was included in the response of one participant. They provided the following response: “Wut? Is this an attention check or a mistake on your part?”. This participant seemed to be confused by our repetition of the premises and conclusion in statement 6. This representation was deliberate. The question “an attention check or a mistake on your part?” indicates that this participant was attentive during this survey as they referred to an authentication test included in our survey meant to determine whether our participants were real humans. Statements 12, 18, 24 and 30 were formulated with the same structure as Statement 6 to test whether participants believe that a fact entails itself, a postulate in Krause et al. (KLM) [1] defeasible logic called reflexivity. However, for this participant, we observe that our representation of reflexive statements was not sufficiently clear. This could be made explicit through the use of additional instruction or alternative syntax. For statement 19, if Nicole A is a cancer patient then Nicole A is terminally ill, the explanation category Other was included in the response of one participant. This participant’s explanation corresponds to category B&Other. This explanation means that the participant endorsed the premises, but not the conclusion and that they provided their own explanation. They provided the following: “Nicole A is a cancer patient, and potentially could be terminally ill”. Again, this explanation suggests a different form of reasoning to the classical logic interpretations of material implication statements. Moreover, it suggests that there are exceptions to the class of cancer patients e.g. cancer patients with non-terminal diagnoses. The willingness to think of exceptions points to reasoning that is consistent with non-monotonic reasoning. For statement 20, if Chris P is a waiter then Chris P has profound knowledge of the menu, the explanation category Other was included in the

response of one participant. This participant's explanation corresponds to the category is A&B&Other. The first part, A, corresponds to endorsing both the premises and the conclusion. The second part, B, corresponds to endorsing the premises, but not the conclusion. This is a contradiction of the first part and is a logical violation. In their response, they provided the following: "He may or may not have profound knowledge of the menu so hopefully this is the right way to answer. It could be made a little more clear how we are supposed to answer in these situations". The participant's choice of explanation categories A and B is consistent with their comment, "He may or may not have profound knowledge of the menu". The comment does not contradict their choice of explanation categories A or B, but confirms it. Yet, the comment also suggests additional information would have persuaded them to endorse only a single explanation. Reasoning in this way would be consistent with belief change. For statement 21, if Daisy M is a crying baby then Daisy M is hungry, the explanation category Other was included in the response of one participant. This participant's explanation corresponds to the B&Other. This means that the participant endorsed the premises, but not the conclusion. This is a logical violation. The response provided was: "There are many reasons a baby like Daisy M could be crying, and being hungry is just a possible answer". This comment contradicts their acceptance of the conclusion in explanation category B. However, the comment, "being hungry is just a possible answer", suggests that there are alternative reasons for crying babies, e.g. some babies cry because they are hungry whilst other babies cry when feeling sick or scared or tired. Without the choice of explanation category B, their reasoning could be viewed as a form of non-monotonic reasoning, where exceptions to the class of typical crying babies exist. For statement 27, if Quentin O is a nurse then Quentin O does have a caring bedside manner, the explanation category Other was included in the response of one participant. The participant provided the following: "I do endorse the statement that Quentin O is a nurse. I mostly endorse the statement that Quentin O does have a caring bedside manner". While the comment does not strictly align with any of the classical logic interpretations of the statement, it does seem to resemble explanation category A in which both the premises and conclusion are endorsed. The comment "mostly" suggests that there may be exceptions to the typical class of caring nurses which aligns with non-monotonic defeasible reasoning. For statement 29, if Rory Z is an accountant then Rory Z is a mathematics boffin, the explanation category Other was included in the response of one participant. This

explanation did not include any of the classical logical interpretations. They provided the following: “I don’t know what a boffin is”. A lack of knowledge on their part is evident by this comment. As a result, their explanation is not complete and cannot be compared to interpretations of the statement in classical logic.

In many cases where participants provided their own explanations, a spectrum of endorsements rather than a positive or negative endorsement would better encompass their beliefs. This confirms what we know about human reasoning: that it is variable, non-monotonic and exception-tolerant. We show the positive ( $\geq 0\%$ ) explanation categories for each statement in the survey in Figures 1–30. This data is provided as the basis for a discussion in future work where the effect of individual statements may be studied for use in a knowledge base for the general English-speaking reasoner. The overall frequency of Other as an explanation, whether as standalone or combined with other explanation categories, accounts for 1,67% of all explanations. Our analysis of the open-ended responses provides scientific evidence in support of the link between human reasoning and non-monotonic reasoning, and between human reasoning and reasoning with belief change. However, it is not significant enough ( $< 50\%$  of all explanations) to influence the analysis of the statements’ plausibility overall.

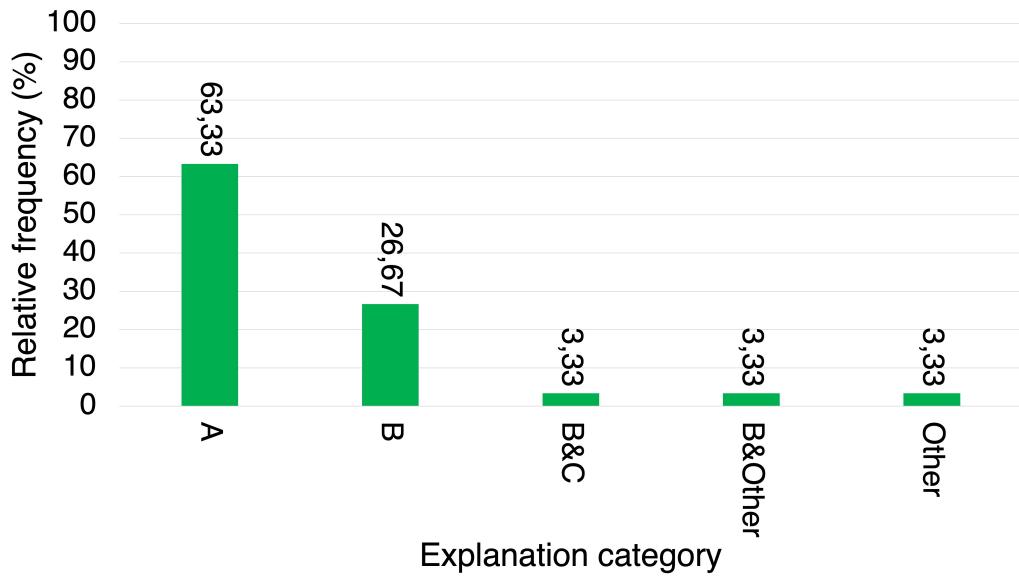


Figure 1: Bar plot of explanation category against relative frequency (%) for statement 1

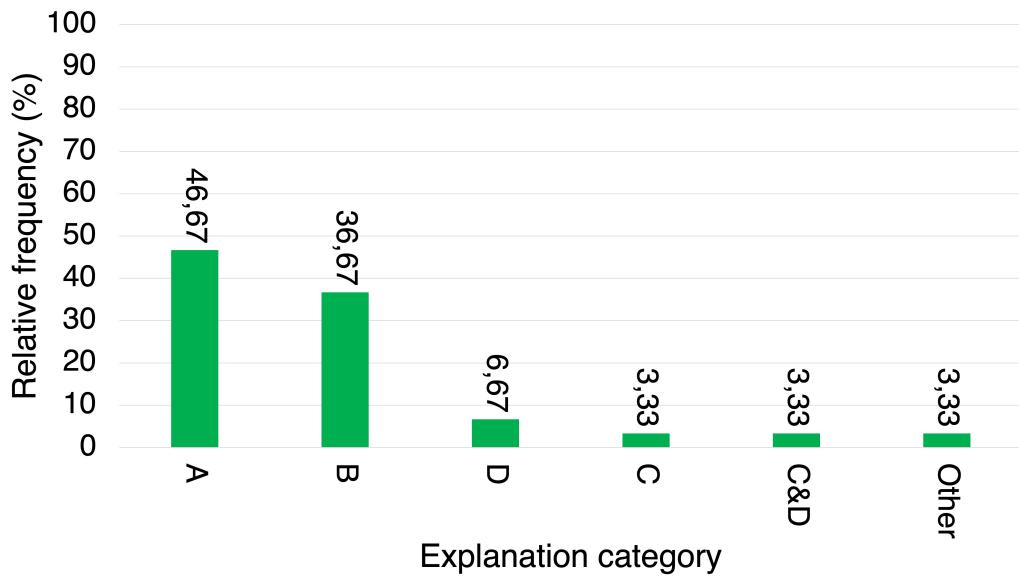


Figure 2: Bar plot of explanation category against relative frequency (%) for statement 2

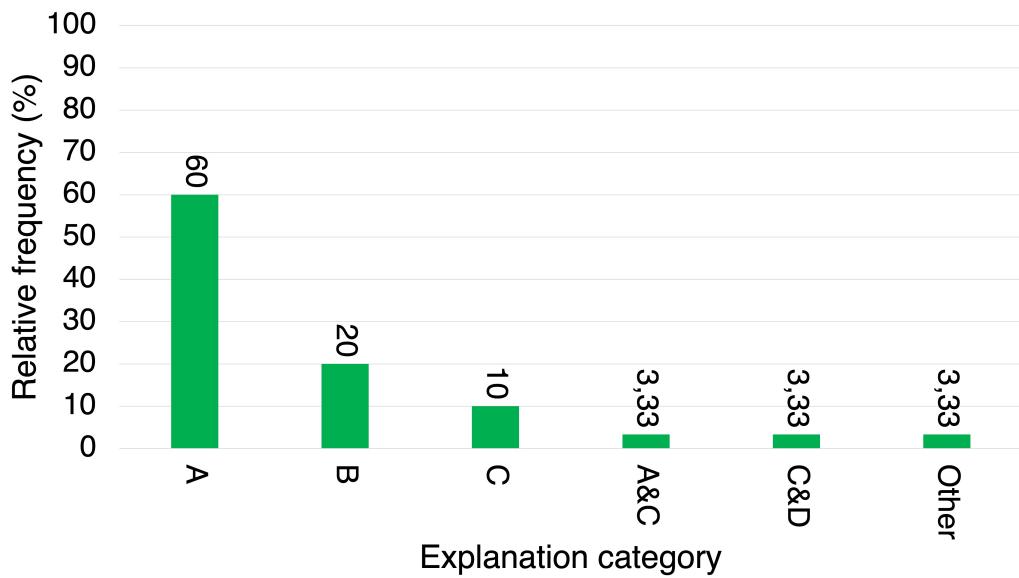


Figure 3: Bar plot of explanation category against relative frequency (%) for statement 3

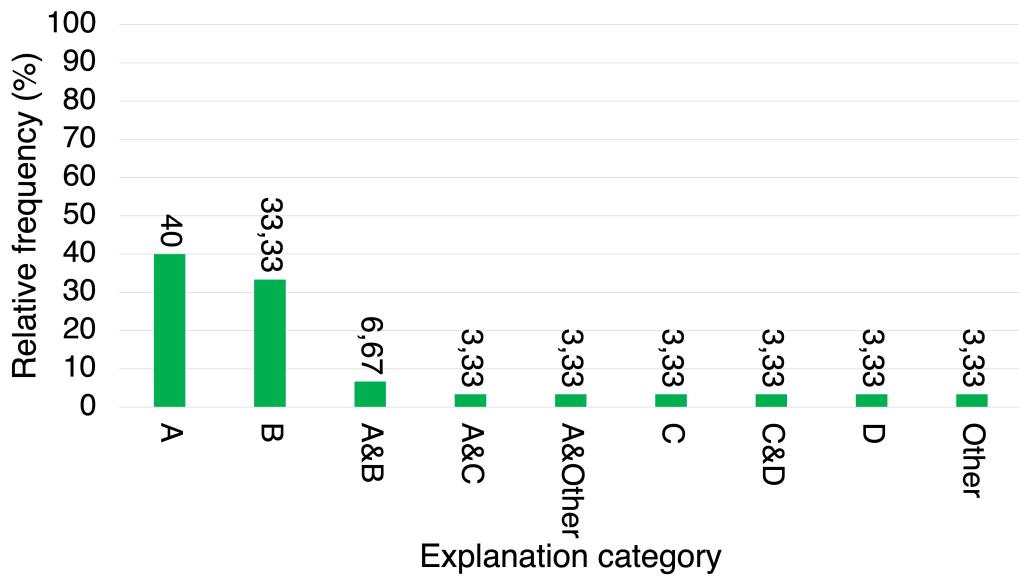


Figure 4: Bar plot of explanation category against relative frequency (%) for statement 4

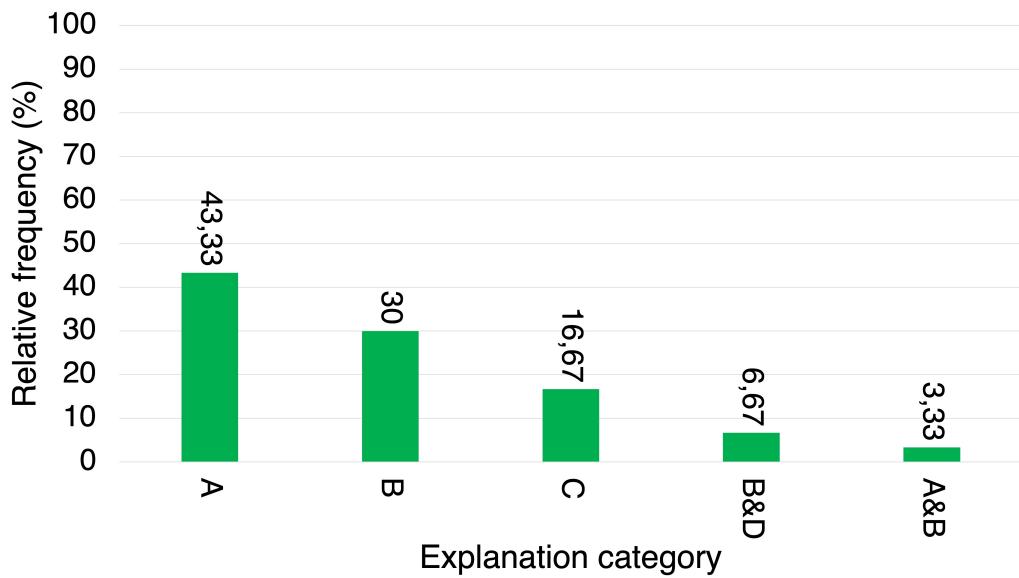


Figure 5: Bar plot of explanation category against relative frequency (%) for statement 5

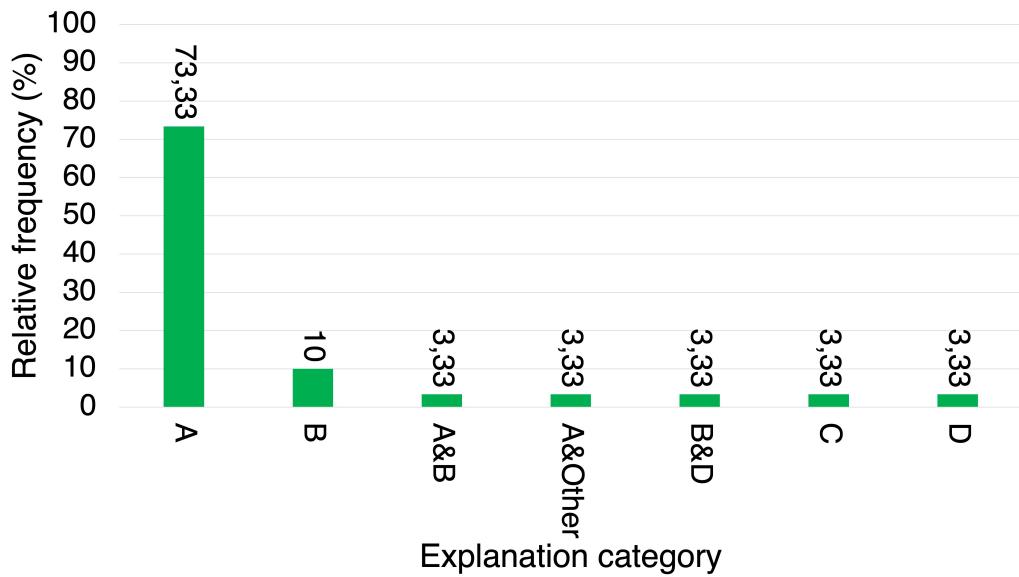


Figure 6: Bar plot of explanation category against relative frequency (%) for statement 6

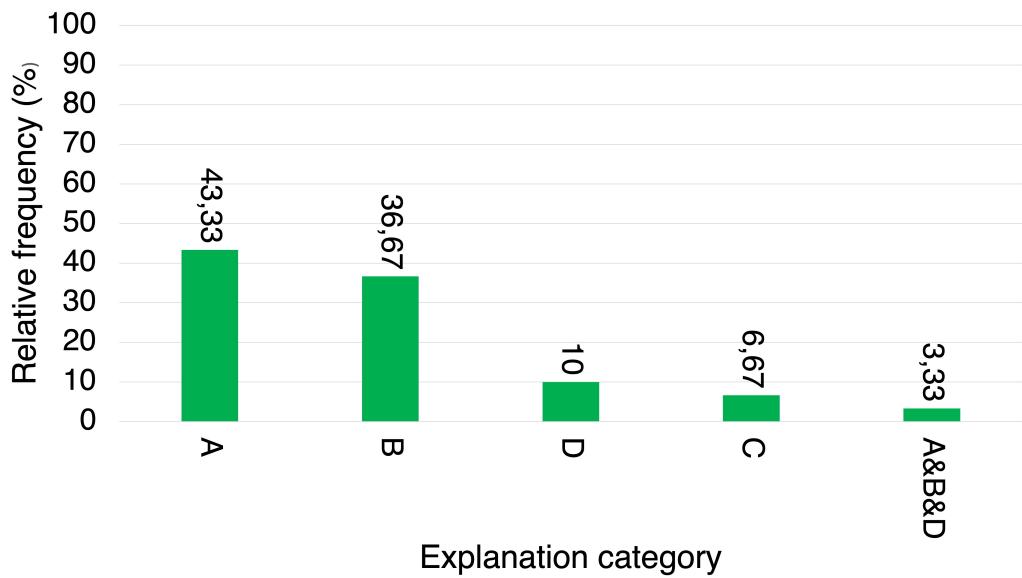


Figure 7: Bar plot of explanation category against relative frequency (%) for statement 7

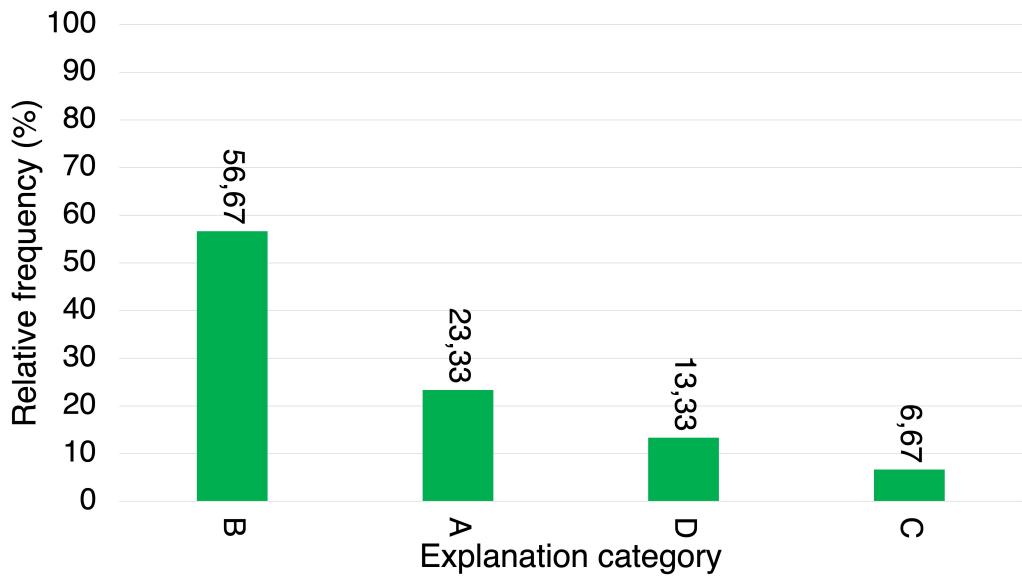


Figure 8: Bar plot of explanation category against relative frequency (%) for statement 8

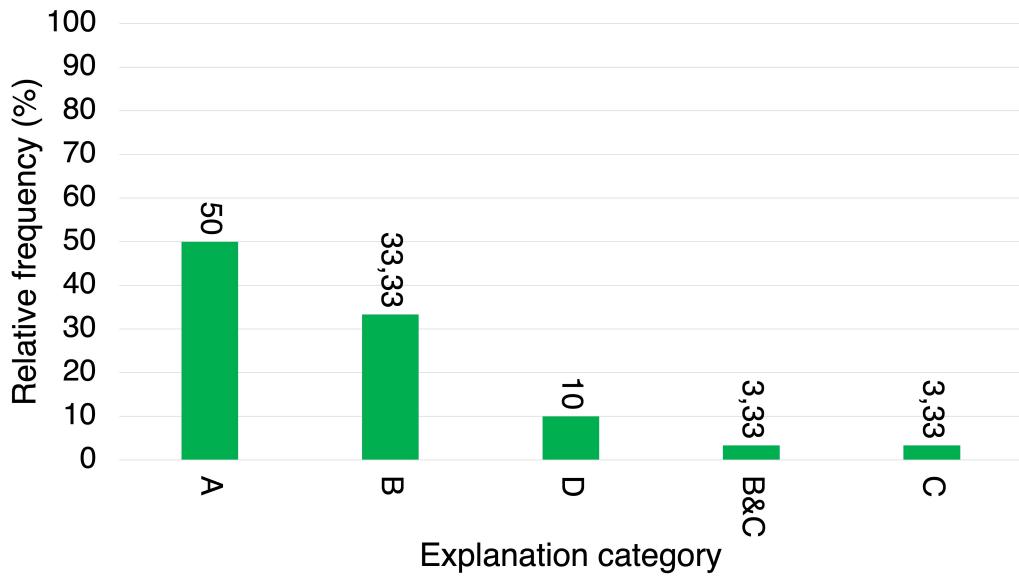


Figure 9: Bar plot of explanation category against relative frequency (%) for statement 9

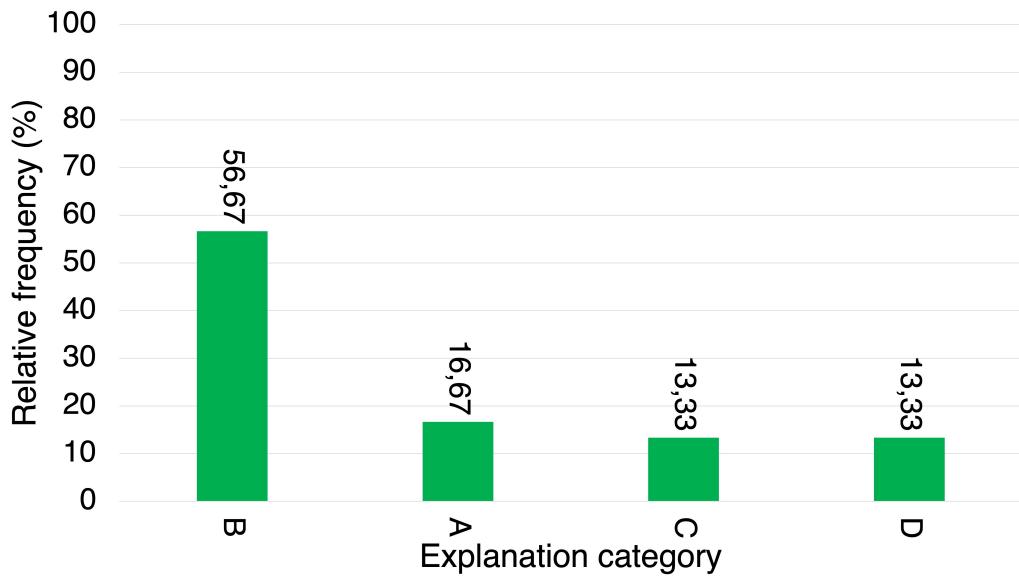


Figure 10: Bar plot of explanation category against relative frequency (%) for statement 10

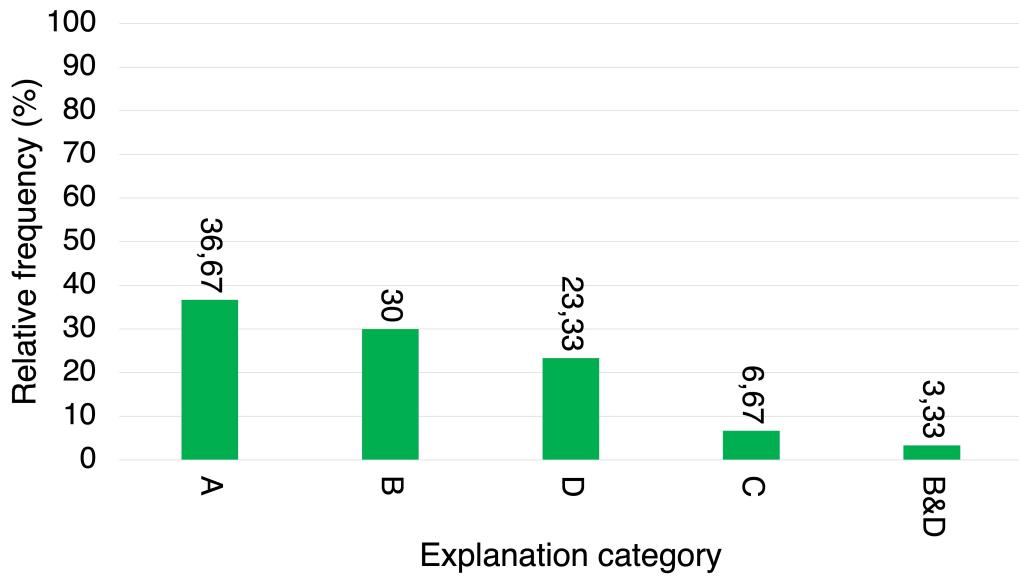


Figure 11: Bar plot of explanation category against relative frequency (%) for statement 11

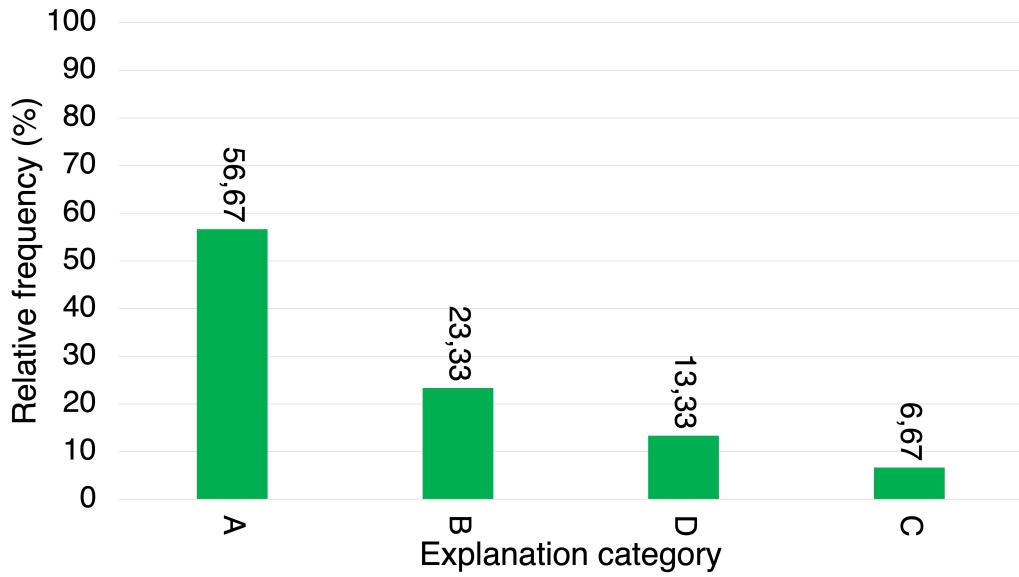


Figure 12: Bar plot of explanation category against relative frequency (%) for statement 12

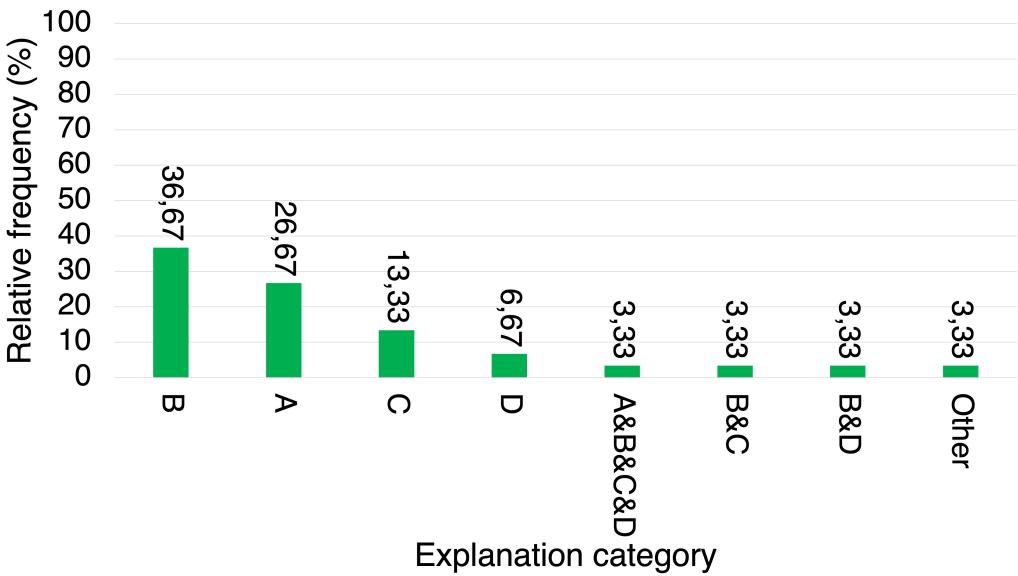


Figure 13: Bar plot of explanation category against relative frequency (%) for statement 13

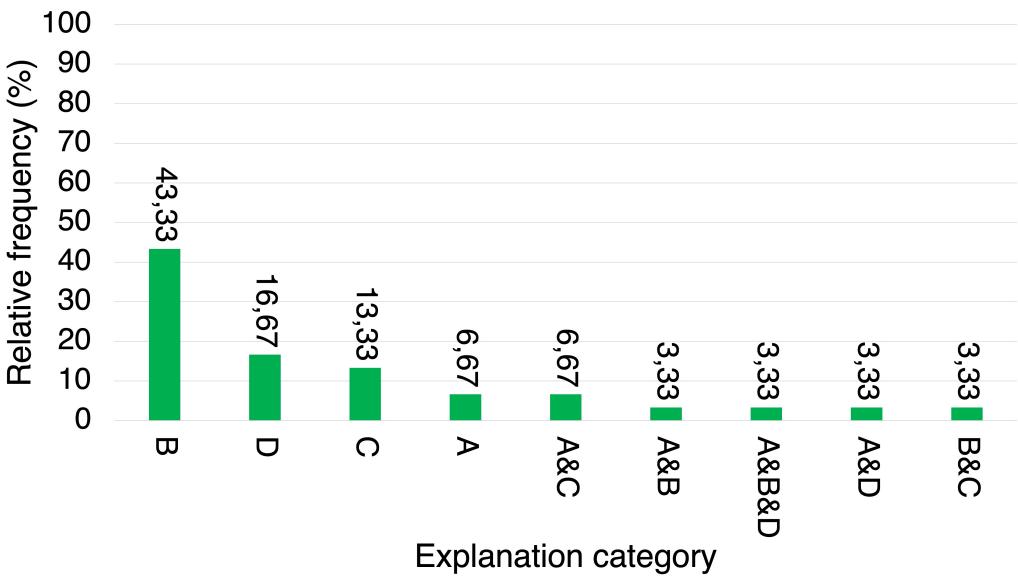


Figure 14: Bar plot of explanation category against relative frequency (%) for statement 14

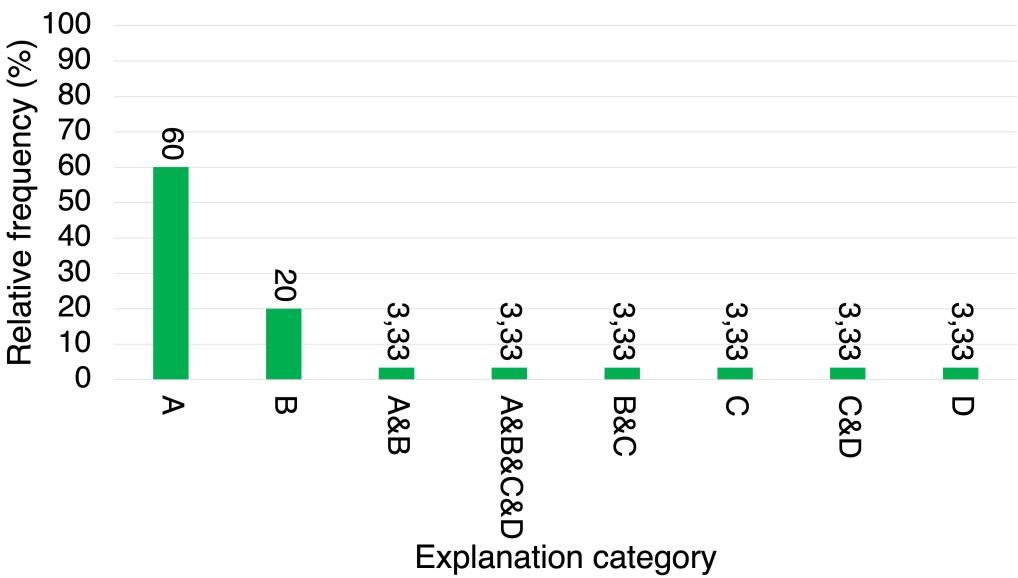


Figure 15: Bar plot of explanation category against relative frequency (%) for statement 15

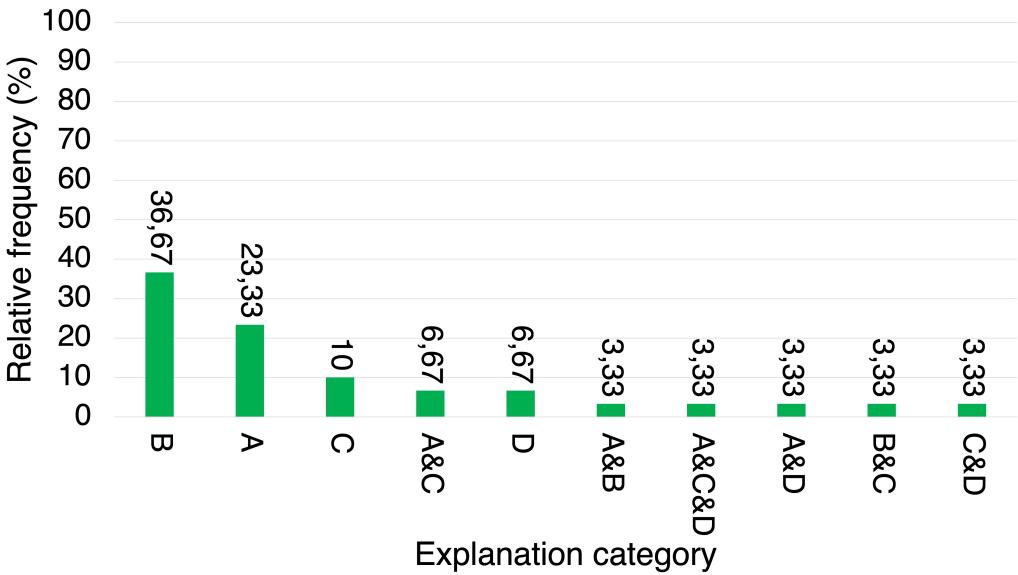


Figure 16: Bar plot of explanation category against relative frequency (%) for statement 16

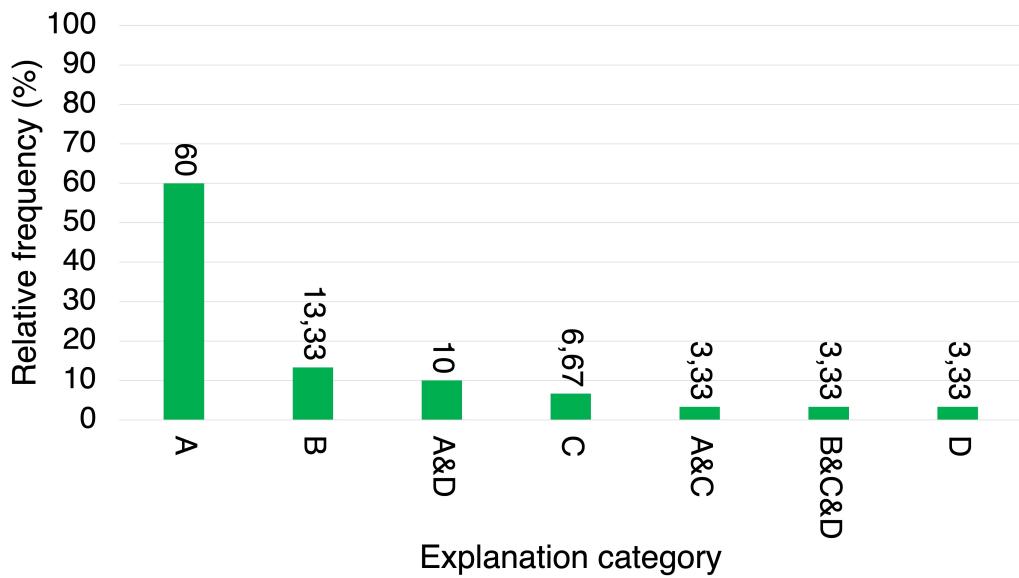


Figure 17: Bar plot of explanation category against relative frequency (%) for statement 17

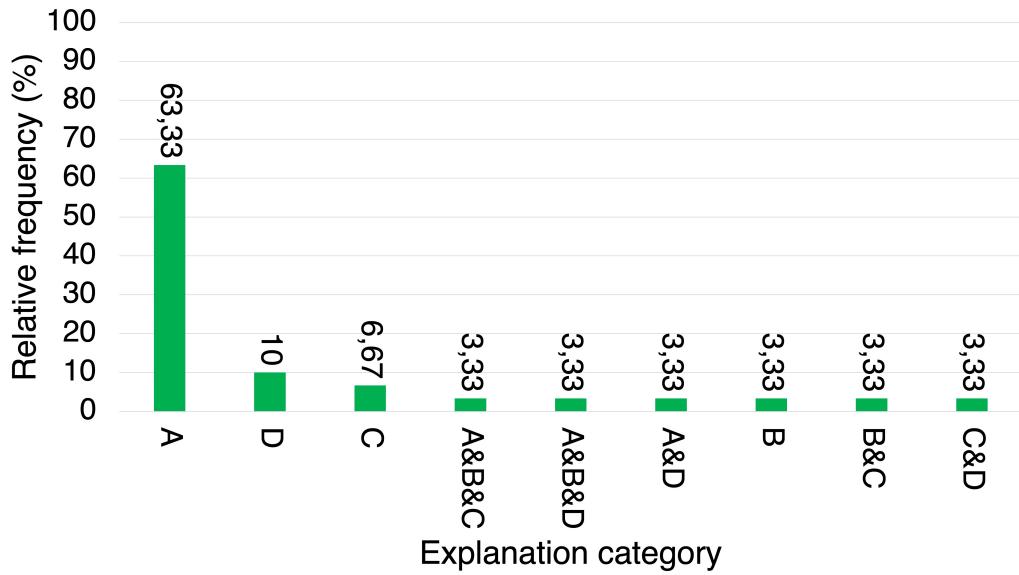


Figure 18: Bar plot of explanation category against relative frequency (%) for statement 18

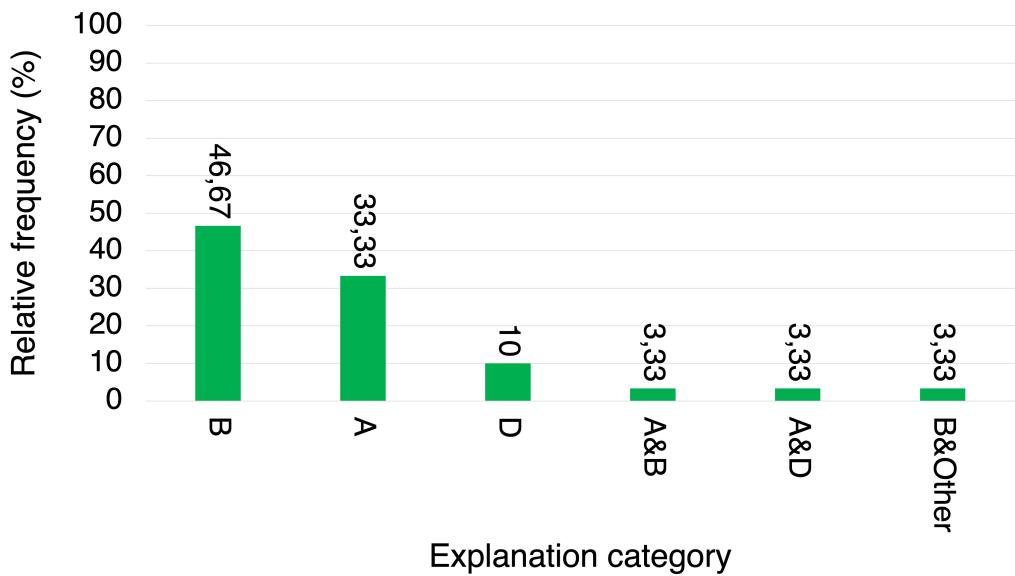


Figure 19: Bar plot of explanation category against relative frequency (%) for statement 19

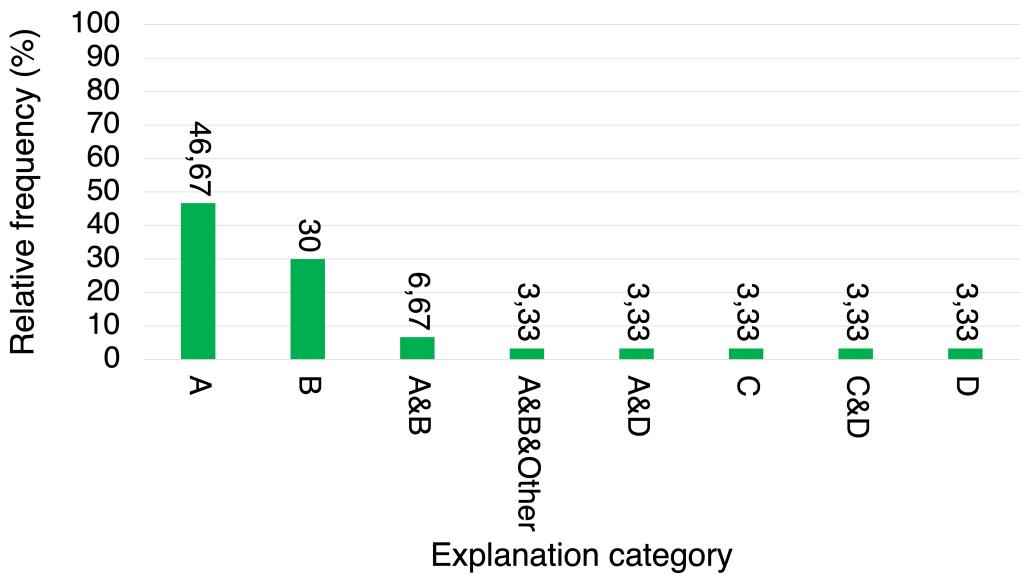


Figure 20: Bar plot of explanation category against relative frequency (%) for statement 20

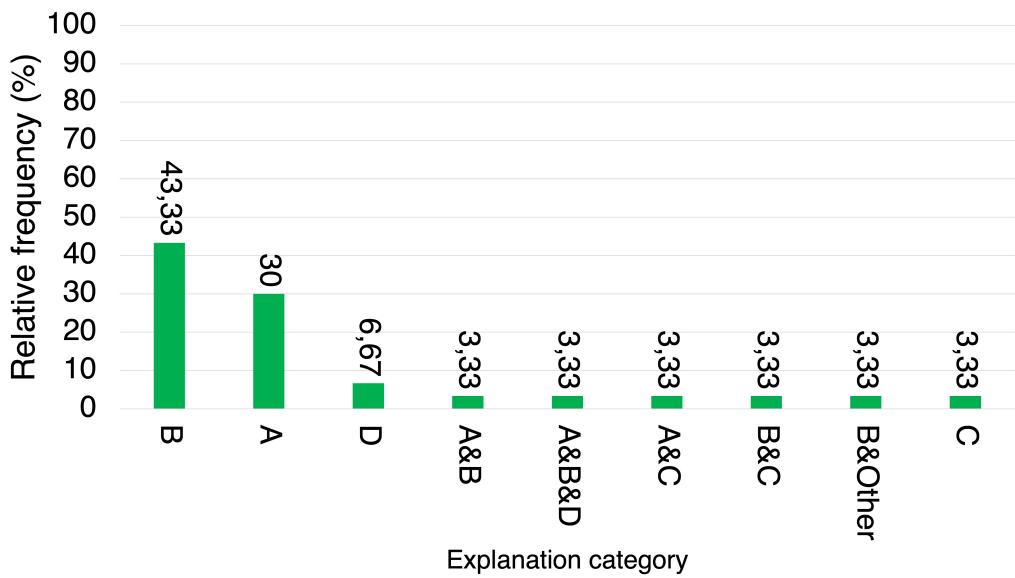


Figure 21: Bar plot of explanation category against relative frequency (%) for statement 21

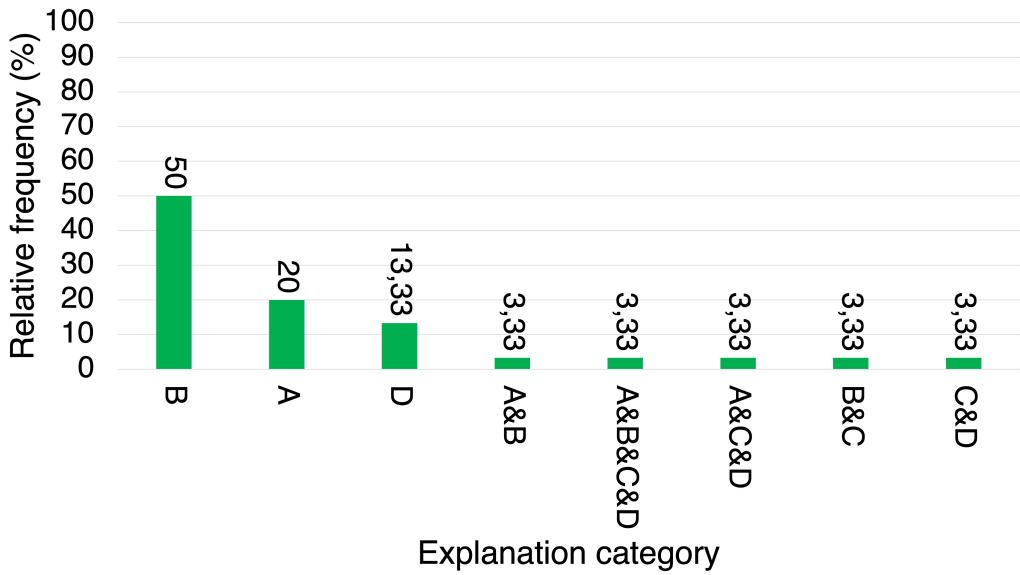


Figure 22: Bar plot of explanation category against relative frequency (%) for statement 22

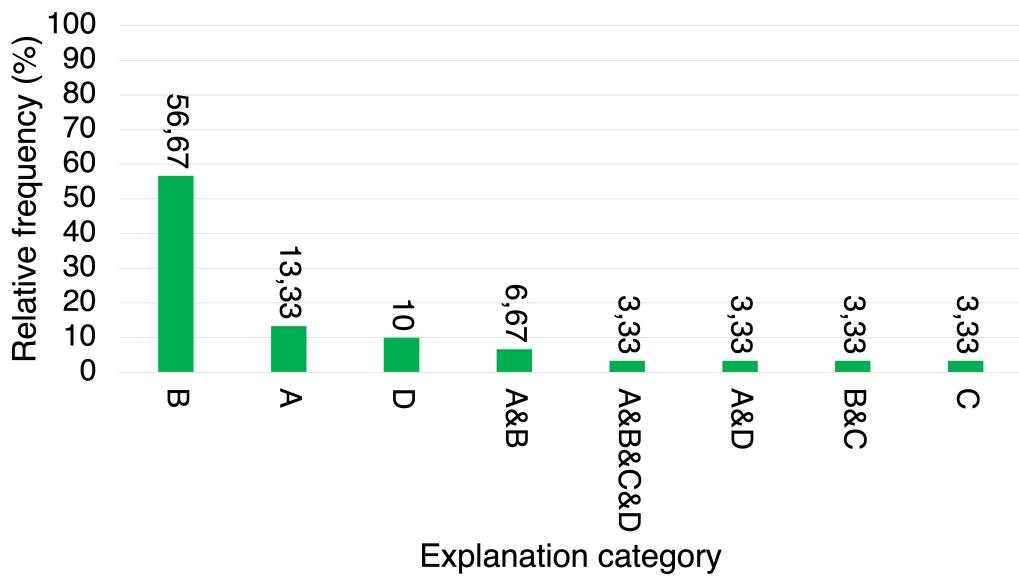


Figure 23: Bar plot of explanation category against relative frequency (%) for statement 23

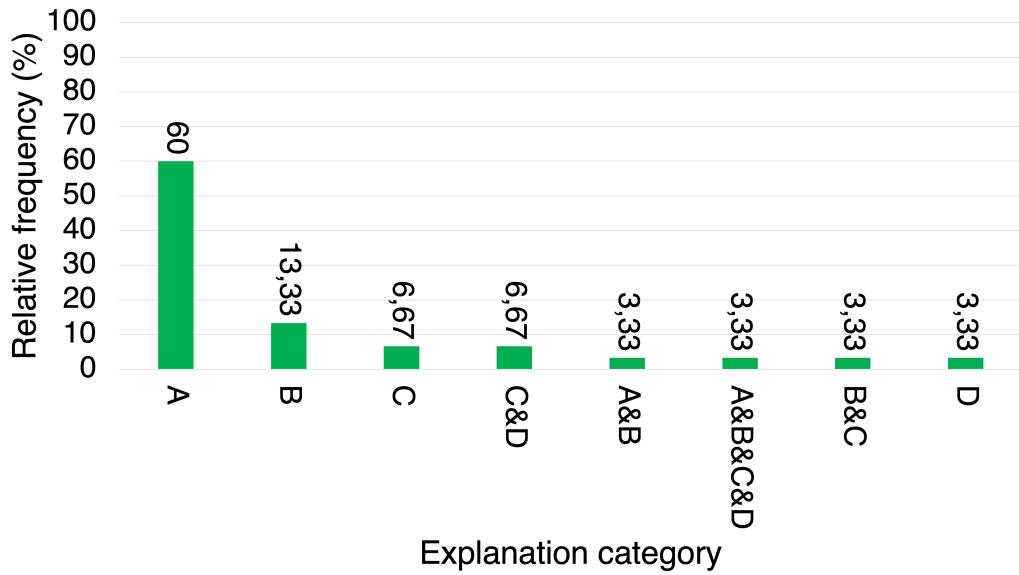


Figure 24: Bar plot of explanation category against relative frequency (%) for statement 24

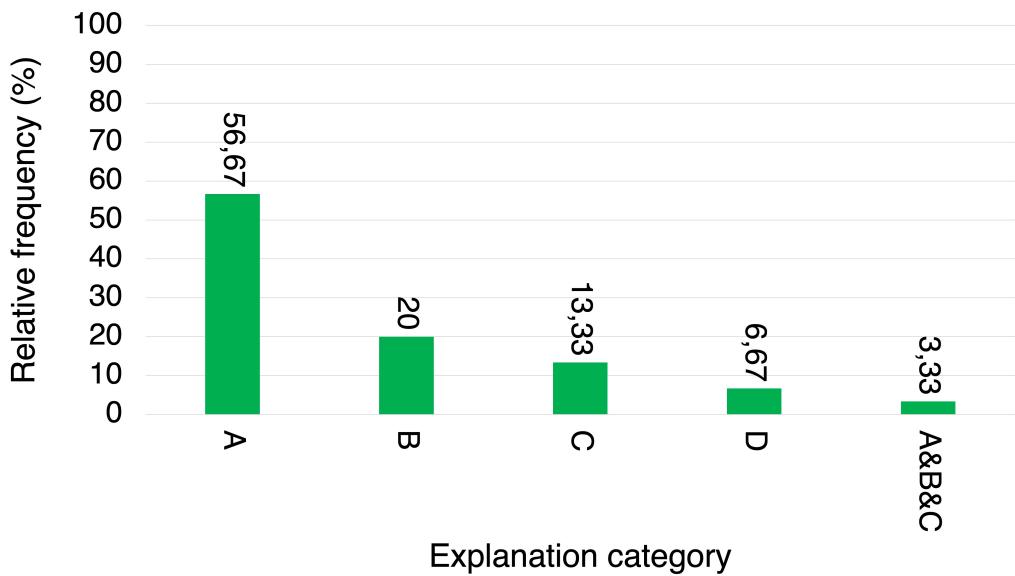


Figure 25: Bar plot of explanation category against relative frequency (%) for statement 25

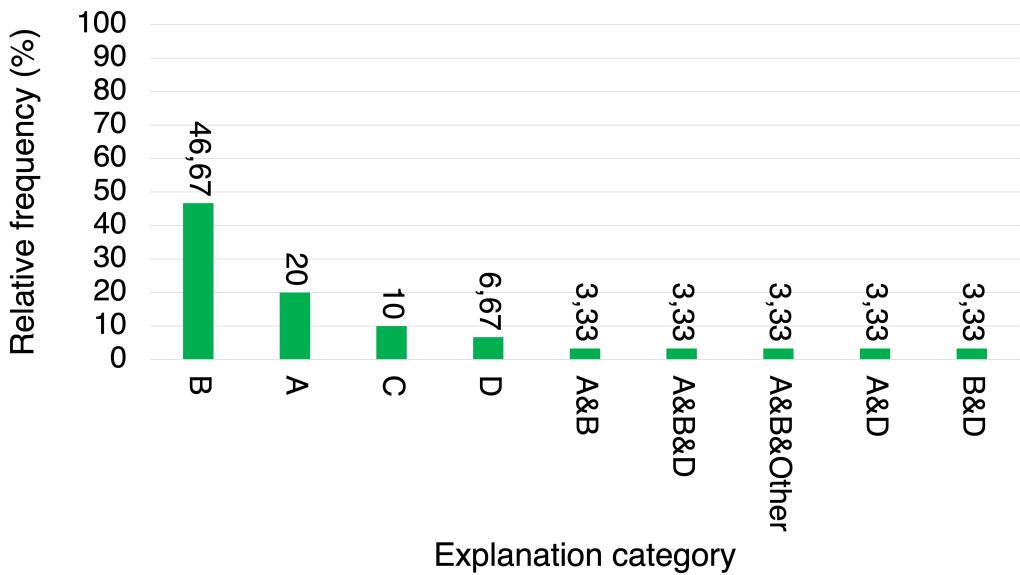


Figure 26: Bar plot of explanation category against relative frequency (%) for statement 26

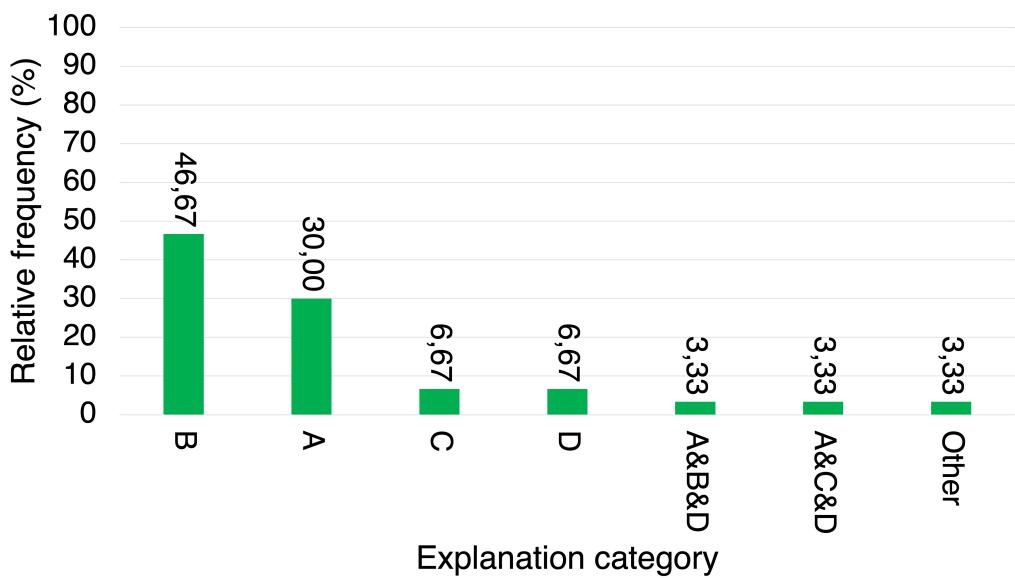


Figure 27: Bar plot of explanation category against relative frequency (%) for statement 27

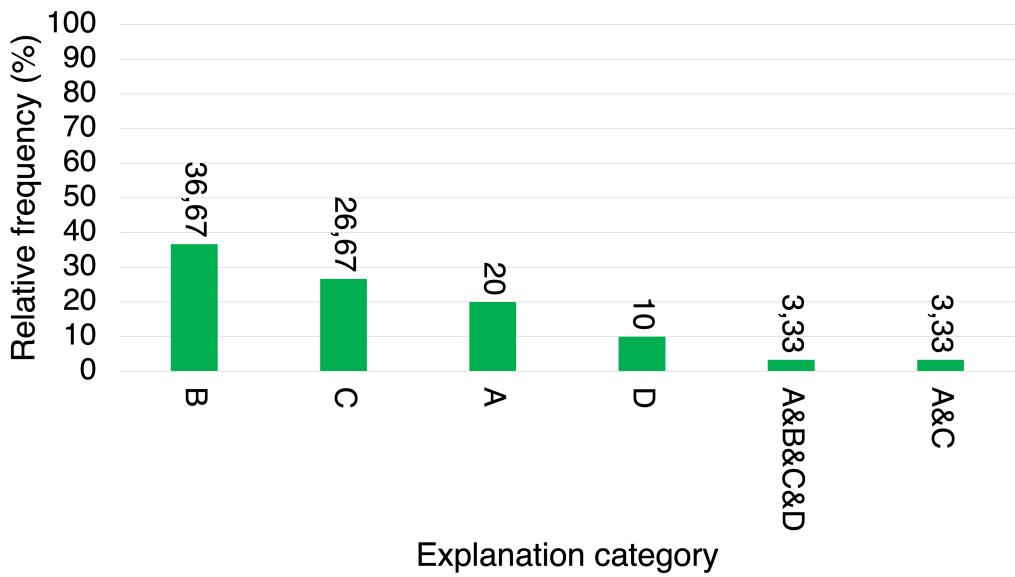


Figure 28: Bar plot of explanation category against relative frequency (%) for statement 28

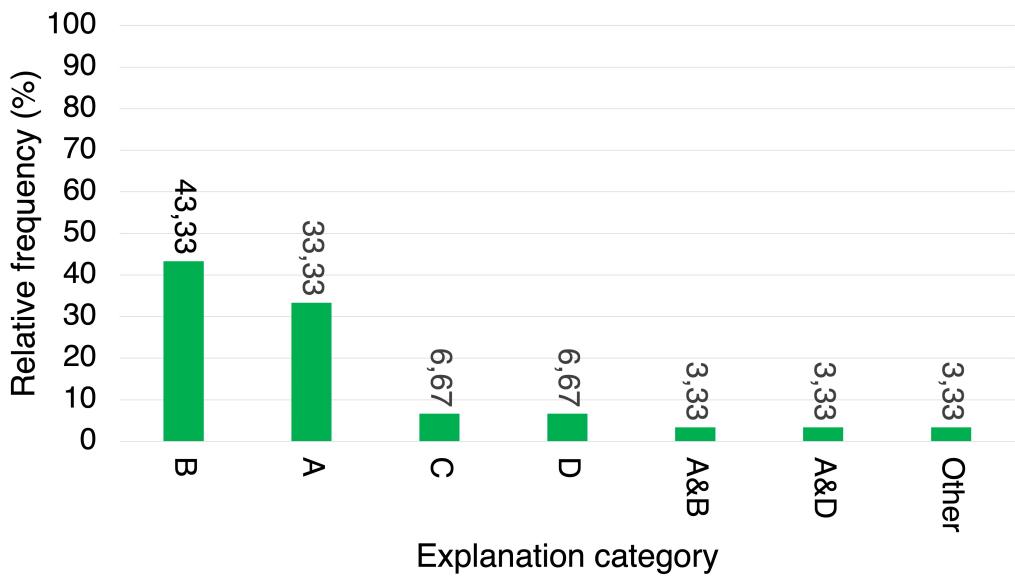


Figure 29: Bar plot of explanation category against relative frequency (%) for statement 29

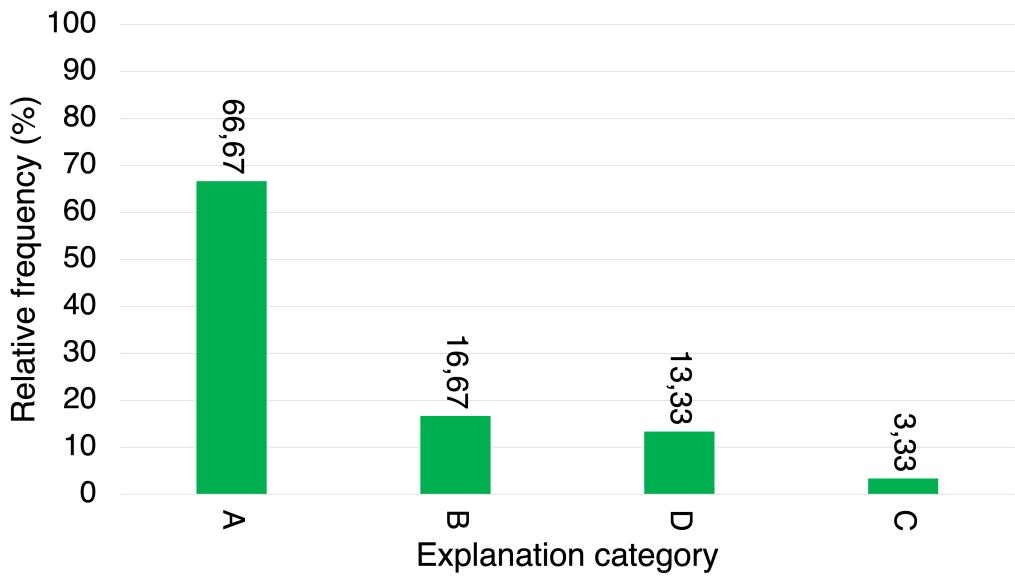


Figure 30: Bar plot of explanation category against relative frequency (%) for statement 30

# Bibliography

- [1] KRAUS, S., LEHMANN, D., AND MAGIDOR, M. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence* 44 (1990), 167–207.