On Floating Conclusions

Daniela Schuster [1]

Universität Konstanz

Jan Broersen [2]

Utrecht University

Henry Prakken [3]

Utrecht University, University of Groningen

Abstract

When there are two lines of argument that contradict each other but still end up with the same conclusion, this conclusion is called a floating conclusion. It is an open topic in skeptical defeasible reasoning if floating conclusions ought to be accepted. Interestingly, the answer seems to be changing for different examples. In this paper, we propose a solution for explaining the different treatments of the floating conclusion in the various examples from the literature. We collect the examples from the literature, extend them with additional examples and test various hypotheses for explaining the difference by means of the examples. We will argue for a framework that accepts a floating conclusion by default but allows for reasons to deviate from the default in order to reject it. These reasons nicely explain the different underlying patterns of our intuitions.

Keywords: Floating Conclusions, Defeasible Reasoning, Skeptical Reasoning

Acknowledgement

Support by the focus area on Human-centered Artificial Intelligence at Utrecht University, Studienstiftung des deutschen Volkes, Baden-Württemberg Stiftung is gratefully acknowledged. We would like to thank the members of the Logic Reading Group at Utrecht University and three anonymous reviewers for their helpful feedback.

[1] daniela.schuster@uni-konstanz.de
[2] j.m.broersen@uu.nl
[3] h.prakken@uu.nl

1 Introduction

Floating conclusions are a phenomenon that appears in the context of defeasible or non-monotonic reasoning. It was investigated already early in [6] and [7]. When there are (at least) two lines of reasoning that contradict each other, but still end up with the same conclusion, this conclusion is called a floating conclusion. One famous example is the Nixon case. In this example, we have one line of reasoning starting from the fact that Nixon is a republican from which it can be concluded (defeasibly) that Nixon is a hawk from which again it can be concluded that Nixon is politically extreme. A different line of reasoning starts from the fact that Nixon is a quaker from which it can be concluded that Nixon is a dove from which again it can be concluded that Nixon is politically extreme. These two lines of reasoning contradict each other, because Nixon cannot be both a hawk and a dove. We have to reject one line of reasoning. Still, both lines of reasoning, albeit conflicting, end up with the same conclusion: the floating conclusion that Nixon is politically extreme. Should we accept this floating conclusion then after all? This is the question that immediately arises and that is going to be the topic of this paper. The name 'floating conclusion' that stems from [7] nicely captures that the conclusion 'floats' above the conflicting arguments. The question of whether we should accept floating conclusions is tied closely to the question of whether we should accept at least one line of reasoning among a set of conflicting lines of reasoning. This builds on the intuition that all the reasoning lines involved have, albeit being fallible, a certain credibility or plausibility. When a conflict between them arises, it becomes clear that at least one line of reasoning fails at some point. Given that it is not clear which line of reasoning fails, we cannot simply accept one and reject the other. However, can we still assume that *there is* (at least) one line of reasoning that is sound? If this is the case, then we should accept a floating conclusion. If this is not the case, then we should not.

Interestingly, there is not one clear answer to this question. In different examples of floating conclusions, we seem to have contradicting intuitions about whether the floating conclusion should be accepted or not. In other words: in conflicting situations, we sometimes think that at least one line of reasoning is sound, while at other times we think that the conflict between the lines of reasoning destroys *both* conflicting reasoning lines. Floating conclusions are one of the most exciting phenomena in the area of defeasible reasoning, but they also pose an unresolved problem in terms of how to deal with them. Therefore, floating conclusions expose possible imitations in defeasible reasoning and also in automated decision-making. A systematic treatment of floating conclusions is missing in the literature so far. Especially in deontic contexts, this can resolve in an alarming inability to derive norms of action in certain situations.

In this paper, we provide a systematic treatment of the phenomenon of floating conclusions. Thereby, we aim to explain the different intuitions concerning the acceptability of floating conclusions in the different examples. It is important to note here that our approach is based on intuitions. We try to provide a theory that manages to explain pre-theoretic intuitions about differ-

ent examples and situations. This method is not undisputed. As it has been noted in [12] and again in [9] the use of intuitions in logic has at least two difficulties. One difficulty is the question of whose intuitions should count (as people might differ in their intuitions). The second difficulty questions the assumption that intuitions should always be taken at face value. In fact, Veltman [12, p. 10] argues that when looking for intuitions, we are usually interested in the pre-theoretic judgments of 'common people' who are no experts in the field and have not been exposed to theories about the topic yet. However, then we cannot tell whether these judgments represent knowledge or barely some kind of fallible belief. Hence, those judgments are fallible and do not provide a "rock bottom empirical basis for testing logical theories" [12, p. 13]. Moreover, (good) theories and arguments can surely guide and change intuitions and pre-theoretic judgments [12, p. 15]. Hence, it is important to not blindly rely on any intuition. Nevertheless, theories should also not contradict all broadly accepted 'common-sense' judgments. Although different people may differ in their intuitions about the acceptability of one floating conclusion or the other, there is clear empirical evidence that some floating conclusions are commonly regarded as acceptable, while others clearly are not (especially when it is considered that people should *act* on these conclusions). A theory that accepts all floating conclusions is just as unsatisfactory as a theory that rejects all floating conclusions. In this paper, we do not blindly rely on any intuition. Rather, as Prakken [9] already suggests, we are searching for some underlying pattern in (commonly shared) intuitions and thereby try to explain similarities and differences.

First, we will present several different examples of non-monotonic arguments that involve a floating conclusion. Many examples are discussed in the literature already, others have been constructed for this paper specifically. We will see that the different examples trigger different intuitions about whether we should accept the respective floating conclusion. Next, we will present different hypotheses that try to explain these conflicting intuitions and we will test the validity of these hypotheses with the help of our examples. After having tested the different hypotheses, we will argue that there is not one single explanation that manages to explain all the different intuitions. Instead, our presented solution will take some ingredients from different explanations. We argue that, per default, floating conclusions are to be accepted. However, there are reasons to deviate from the default and to reject a floating conclusion. We will present two different reasons for deviation that together nicely explain and cover all the presented examples. One reason applies if there is a possible 'compromise' between the conflicting elements of the arguments; the second reason applies if the conflict is harmful not only to the conflicting part of the argument but also to other non-conflicting parts because the conflict undermines the credibility of the sources of information altogether. Both of these reasons are based on the fact that in situations of floating conclusions, the conflicting propositions

are *contrary*.[4] This means that the propositions cannot be true together, but yet can be false together. The two explanations which give us reason to deviate from the default both spell out a way in which the conflicting propositions are both false, offering a third alternative beyond the two (in the arguments displayed) alternatives that one proposition is true and the other one false or vice versa.

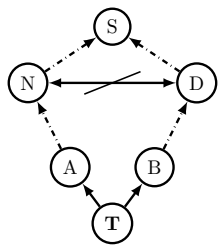
2 Examples of floating conclusions

We will present examples of arguments that involve a floating conclusion. Some of the examples can be found in the literature, others are invented for this paper in order to obtain a precise impression of the phenomenon that is as comprehensive as possible. In a second step, we will divide them by means of the different intuitions about the acceptance of the respective floating conclusion.

2.1 Presenting the examples

In the following, we will use capital letters to abbreviate the sentences or propositions. The arguments are visualized via arrows connecting the sentences. The non-dashed arrows represent strict, monotonic reasoning, while the dashed arrows represent defeasible inferences.[5] The double-sided crossed-out arrow visualizes a conflict between two sentences, while the **T** stands for 'truth.' Sentences that follow from **T** are taken to be known.

Ice-Skating [9]



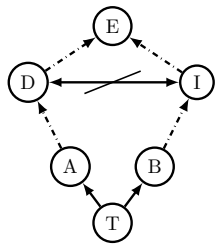
A Brigt Rykkje has a Norwegian name.
B Brigt Rykkje was born in the Netherlands.
N Brigt Rykkje is Norwegian.
D Brigt Rykkje is Dutch.
S Brigt Rykkje likes ice-skating.

The argument that is visualized by the picture then reads as follows: It is both true (hence strictly follows from (**T**)) that Brigt Rykkje was born in the Netherlands (B) and that he has a Norwegian name (A). The argument on the right side tells us that Brigt Rykkje is Dutch (D) since he was born in the Netherlands (B). The argument on the left tells us that Brigt Rykkje is Norwegian (N) since he has a Norwegian name (A). These two statements, (N) and (D), however, contradict each other and cannot be both true at the same

[4] Thanks to Michael De for pointing us towards this.
[5] One could also call them material inferences in the terminology of [11] and [1].

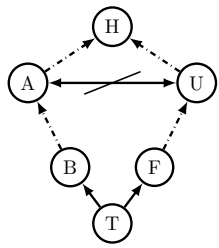
time.[6] On the right side, the argument continues: Brigt Rykkje likes ice-skating, since he is Dutch (D). On the left side, the argument continues: Brigt Rykkje likes ice-skating (S), since he is Norwegian (N). Hence, both argument lines end up with the floating conclusion that Brigt Rykkje likes ice-skating (S). All reasoning steps here are beliefs.

Economy [4]



A Economist A says we will have deflation.
B Economist B says we will have inflation.
D We will have deflation.
I We will have inflation
E We will have economic downturn.

Student Housing [2]

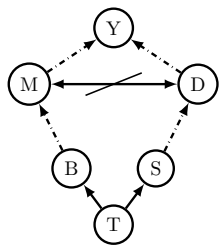


B Susan's boyfriend studies in Amsterdam.
F Susan's best friend studies in Utrecht.
A Susan wants to study in Amsterdam.
U Susan wants to study in Utrecht.
H It will be difficult to get housing.

Here not only beliefs but also desires are involved. For example: Susan *wants* to study in Amsterdam (A), because her boyfriend studies in Amsterdam (B). But again, she *believes* that housing will be very expensive if she studies in Amsterdam.

[6] In almost all the examples, we are making some empirical assumptions, like here: It is not possible to have two citizenships. This reflects the fact that we are reasoning in a non-monotonic setting with incomplete knowledge. One could say that the inferences we draw are material inferences rather than formal inferences in the terms of Sellars [11] and Brandom [1].

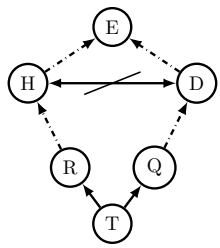
Yacht [4]



B My brother tells me that dad will give his half a million dollars to him, but mom will give it to me.
S My sister tells me that mom will give her half a million dollars to her, but dad will give it to me.
M I will get half a million dollars from my mom.
D I will get half a million dollars from my dad.
Y I put a high deposit on a Yacht that costs half a million dollars.

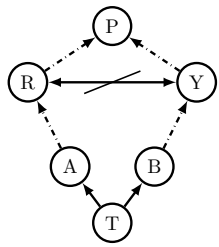
This example from Horty [4] is about a situation where I have a brother and sister. Our parents are separated and will both die soon. The parents each have a fortune of half a million dollars. Before both parents went into comas, my brother talked to my father and my sister talked to my mother. My sister tells me that according to my mother, my mother will give her half a million dollars to her (my sister), but my father will give his half a million dollars to me. My brother tells me that according to my father, my father will give his half a million dollars to him (my brother), but my mother will give her half a million dollars to me. In this story, I really want to buy a (very particular) yacht for half a million dollars and I intend to make a very large down payment on the yacht should I receive half a million after my parents die.

Nixon [4]



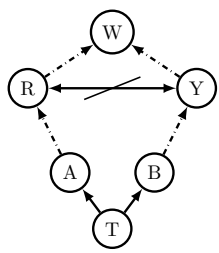
R Nixon is a republican.
Q Nixon is a quaker.
H Nixon is a hawk.
D Nixon is a dove.
E Nixon is politically extreme.

Primary Color



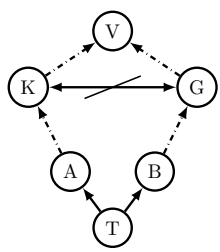
A Anna says that the cup is red.
B Ben says that the cup is yellow.
R The cup is red.
Y The cup is yellow.
P The cup is colored in a primary color.

Wavelength Color



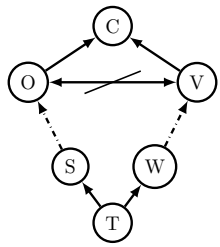
A Anna says that the cup is red.
B Ben says that the cup is yellow.
R The cup is red.
Y The cup is yellow.
W The color of the cup has a higher wavelength than the wavelength of blue.

Murderer [9]



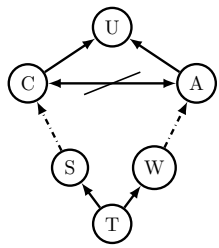
A Witness A says that Peter killed the victim with a knife.
B Witness B says that Peter killed the victim with a gun.
K Peter killed the victim with a knife.
G Peter killed the victim with a gun.
V Peter killed the victim.

Mary in Canada [10] and [4]



S Mary's spouse lives in Toronto.
W Mary's work is in Vancouver.
O Mary lives in Toronto.
V Mary lives in Vancouver.
C Mary lives in Toronto or in Vancouver.

Carol in the US [4]



S Carol's spouse lives in College Park.
W Carol's work is in Alexandria.
C Carol lives in College Park.
A Carol lives in Alexandria.
U Carol lives in College Park or in Alexandria.

2.2 Intuitions about the acceptance of the floating conclusions in the examples

Floating conclusion accepted	Floating conclusion rejected
Ice-Skating	**Economy**
Student Housing	**Yacht**
Wavelength Color	**Primary Color**
Mary in Canada	**Carol in the US**
	Nixon
	Murderer

The table shows the examples in which the floating conclusion should be intuitively accepted and the examples in which it should not. As mentioned in the

introduction, the use of intuitions raises some questions. Of course, some people's intuitions may diverge from the table. Although we have not conducted a scientific study, we have asked enough people about their intuitions regarding these examples to assume that the table is representative. Of course, we are also aware that most of the people we interviewed come from an academic background and that the intuitions of other groups of people might be different.

The Nixon case is probably the most controversial and therefore the most interesting case. In former literature [3] people argued that the floating conclusion in the Nixon case should be accepted. However, we think that it should be rejected. Especially when we explain our reasons *why* the Nixon floating conclusion should be rejected, people seem to sometimes change their intuitions and admit that in fact, one cannot conclude that Nixon is politically extreme. Without presuming additional knowledge about Nixon as a person, it is natural to think that his quaker and his republican side 'balance each other out' such that he ends up with no politically extreme stance. This reflects nicely what we said in the introduction. Intuitions are not infallible and they are not necessarily stable. Sometimes good explanations can change intuitions.

3 Hypotheses

In this section, we present different hypotheses that aim to explain why some but not all floating conclusions seem acceptable. We do not claim that the list of hypotheses is exhaustive, nor that all of the hypotheses are prima facie equally convincing. The list contains the hypotheses we found in the literature so far and new hypotheses that we took to be reasonable and worth mentioning. In the subsequent subsection, we will then evaluate the hypotheses by virtue of our examples.

3.1 Presenting the hypotheses

(i) **Vagueness**: One possible explanation is bound to the concept of vagueness. Some conflicts can be seen as borderline cases for vague concepts that are involved in the corresponding defaults. If a vague concept is involved, and the conflicting propositions incorporate a clear, non-borderline case of the concept, it has to be tested whether the floating conclusion also follows from the borderline case. If the floating conclusion does not follow from the borderline case, it should be rejected.

(ii) **The direction of fit** [2]: The difference could stem from different direction of fits. Beliefs are propositions that aim to describe the world, hence the direction of fit can be described as proposition-to-world. Desires and intentions, on the other hand, are propositions that describe how the world ought to be, so the direction of fit is world-to-proposition. This is why conflicting beliefs 'cancel each other out,' resulting in the rejection of the floating conclusion. Conflicting desires or intentions, on the other hand, do not cancel each other out. Thus, at least one of the desires will remain intact and the floating conclusion is to be accepted.

(iii) **Hidden Defaults** [9]: This explanation states that the reason that some

floating conclusions might seem unacceptable, results from implicit "hidden defaults" that are not mentioned explicitly, but have to be thought along the respective examples. These hidden defaults defeat (through undercutting) the presented defaults that lead to the alleged floating conclusions, such that these in fact are no floating conclusions but conclusions of defeated defaults.

(iv) **Possible Compromise**: This explanation suggests that one has to look at the compatibility of the conflicting propositions. If there is a possible 'compromise,' or intermediate position, between the conflicting propositions, it is likely that this compromising position is in fact the case. In such a situation, one has to check if the floating conclusion also follows from the compromising case. If it follows *only* from the presented 'extreme' cases but not from the compromising one, the floating conclusion must be rejected. If there is no compromise between the conflicting propositions, it is justified to think that at least one of the conflicted propositions is true, and hence that the floating conclusion is acceptable.

(v) **Harmfulness of the conflict**: This explanation takes a closer look at the conflict, as well as at the sources of information. Sometimes it seems that the conflict is only harmful to the conflict itself. In other cases, though, the conflict seems to destroy the credibility of the sources of information more generally. If this is the case, there is no longer a reason to assume that at least one line of reasoning is sound which results in rejecting the floating conclusion.

3.2 Testing the hypotheses

In this section, we will test the presented hypotheses by means of our examples. We will see that, while most hypotheses manage to explain certain examples well, no hypothesis manages to explain the intuitions behind every example presented.

(i) **Vagueness:** The vagueness hypothesis is motivated by examples like **Wavelength Color** and **Primary Color**. These two examples involve a vague concept (a color). Clearly, the conflicting propositions (that the cup is red and that the cup is yellow) can be dissolved by a third proposition (that the cup is orange) representing the borderline case. In the **Primary Color** example, the floating conclusion does not follow from the borderline case (thus the conclusion is rejected), for **Wavelength Color** the floating conclusion does follow from the borderline case as well (thus it is accepted). The involvement of vagueness alone does not do the job of explaining the differences though. Moreover, there are plenty of examples that do not involve vagueness and for which we still have varying intuitions. These cannot be explained by this hypothesis. However, it becomes clear quite easily that vagueness alone cannot explain all examples. For example, there is **Yacht** which is a rejection example and **Ice-Skating** which is an acceptance example, but neither of the examples involves a vague concept.

(ii) **The direction of fit:** The idea that a different direction of fit can lead to different intuitions about the acceptability of floating conclusions was originally motivated in [2] by the different intuitions in the examples **Economy** and **Student Housing**. In the latter example, the conflict arises due to conflicting desires. Susan wants to study both in Utrecht and in Amsterdam. Although it is clear, that one desire will 'defeat' the other eventually, the desires do not cancel each other out as in the Economy case where we have a conflict between beliefs. However, this explanation fails in other examples. **Ice-Skating** is an example that is free of desires and intentions and purely based on beliefs. Still, we want to accept the floating conclusion in **Ice-Skating**. [7]

(iii) **Hidden Defaults:** [9] argues that the examples **Yacht** and **Murderer** do in fact not provide a reason to reject floating conclusions. The propositions that conflict each other and from which the floating conclusions follow are in both cases defeated since the defaults leading to these conclusions are undercut by other defaults, that are not mentioned explicitly in the theory. In the case of **Murderer**, what makes the alleged floating conclusion unacceptable is the hidden default that, if two witnesses say contradicting things, their credibility is dismissed. This default then undercuts both the default that concludes that Peter killed the victim with a gun and the default that concludes that Peter killed the victim with a knife, yielding no floating conclusion. Likewise in the **Yacht** example, a hidden default will undercut both arguments that rely on the testimonies of my sister and my brother. [8] This strategy succeeds in other examples as well. In the Nixon case, one could find an additional, hidden default stating that if someone is both a quaker and a republican, one cannot tell anything about his or her opinion with respect to military operations. This hidden default would then undercut both defaults that infer either that Nixon is a dove or that Nixon is a hawk. The floating conclusion that he is politically extreme would then not follow either. The rather clear case of **Ice-Skating** also speaks in favor of this hypothesis. There is no apparent hidden default that should be visualized in the example, leading to the intuitively correct conclusion that the floating conclusion is acceptable. The strategy, however, becomes more questionable when examples like **Mary in Canada** and **Carol in the US** (or the Color examples)

[7] One can easily see how the explanation fails in the other direction as well. If one adapted **Carol in the US** to an example about Carol's desires to live in one and the other city, the hypothesis would state that the floating conclusion is to be accepted, although we want to reject it.

[8] Note that Prakken [9] described the example slightly differently. In his description, both my sister and my brother tell me that they spoke to both parents and that my mom (respectively my dad) told my sister (respectively my brother) that she will give me her (his) money. Prakken argues that this example relies on the additional default that people tend to speak the truth about their intentions, which is undercut as soon as people (in this case both mom and dad) tell conflicting things about their intentions.

are considered, where the same defaults in one case lead to seemingly acceptable floating conclusions and in another case to unacceptable ones. Why should there be hidden defaults in one case but not in the other? Prakken himself also admits that this strategy might not be valid for all possible examples, such as conflicts due to different interpretations of legal norms. Moreover, we think that, although this thesis might be applicable for a lot of examples, it does not really provide an explanation about *why* in certain situations a floating conclusion is to be accepted and in others not. By referring only to possible missing defaults, we might get a way out of the unequal treatment of the different floating conclusions, but it still shifts the burden of explanation only to the question about why we feel like there are some defaults missing (or hidden) in some cases, while in other cases this is not so.

(iv) **Possible Compromise:** The idea behind this thesis can best be visualized by the different intuitions of the **Mary in Canada** and **Carol in the US** case. Although the defaults leading to the conflict and to the floating conclusions are of the exact same form, the floating conclusion seems justified in one case and not in the other (as [4] notices.) What explains the difference in this particular case? It seems like the conclusion that Mary lives either in Vancouver or in Toronto is acceptable because there is not really an alternative option in the 'middle.' Since the cities are extremely far away from each other, it is not likely that Mary could live somewhere in the middle and commute between the places on a daily basis. This is different in the case of Carol in the US. Since both cities, College Park and Alexandria, are in fact not very far away from each other and there is a good 'compromise,' Washington D.C., that is in the middle, it is likely that Carol neither lives in College Park nor in Alexandria, but went for the compromise, the city in between. This idea can be transferred to other examples, too. In the **Economy** case, there is a 'compromise'[9] between (strong) deflation and inflation, namely that there will be none of both. Likewise in the **Nixon** case, the compromise between Nixon being a Hawk and Nixon being a Dove lies clearly in the middle in describing Nixon as not having a clear or extreme opinion on military use. In both cases, we do not want to accept the floating conclusion, because the compromise is just too likely and from the compromise the floating conclusion does

[9] Note that the use of the word 'compromise' may be somewhat unusual in this context. Not in all the cases described is there really a compromise in the sense of people agreeing on something. What we mean here by compromise is rather an unignorable possibility or relevant alternative. We use the word 'compromise' anyway because it suggests so nicely that this alternative or possibility lies somewhere *in the middle* on a spectrum at the end of which the two conflicting options lie (and is not simply some additional alternative that lies outside the spectrum considered so far).

not follow.[10] This is different in the cases of **Ice-Skating** and **Student Housing**. There is no attractive student town between Amsterdam and Utrecht and even less is it possible, that Brigt Rykkje can have a citizenship 'in between' Norwegian and Dutch. Therefore, we should stick with the conclusion that, even if he cannot have both, he has at least one of the citizenships, such that the floating conclusion can be drawn. The general idea is that if there is no compromise between the conflicting propositions, then it is likely that at least one line of reasoning is correct and the floating conclusion will follow. If there is a plausible compromise, then it has to be tested if the floating conclusion follows from this compromise, too. This can be nicely visualized via the two color examples. In the identically constructed examples **Wavelength Color** and **Primary Color**, the compromise (that the cup is orange) entails one floating conclusion (that the cup is colored with a higher wavelength than the one of blue) but not the other floating conclusion (that the cup is colored in a primary color). However, the examples of **Murderer** and **Yacht** cannot be perfectly explained by this hypothesis. The reason why we want to reject the floating conclusion is not that there seems to be a compromise or intermediate position between the two conflicting propositions. Rather, it seems like the sole fact that there *is* a conflict undermines the credibility of both argument lines.

(v) **Harmfulness of the conflict:** The thesis about the harmfulness of the conflict is based exactly on this observation concerning **Yacht** and **Murderer**. The basic idea is that there are different kinds of conflicts. Some kinds of conflicts are harmful to the floating conclusion, others are not. The cases of **Yacht** and **Murder**, for example, involve a conflict in which two witnesses assess different things that, although in conflict with each other in some respect, are consistent with each other in another respect. In the Murderer case the witnesses' testimonies conflict in respect to the murder weapon they describe Peter to have used, but they agree upon the fact that it was Peter who killed the victim. In the case of the Yacht, the siblings' testimonies are in conflict with each other in respect to what Mom and Dad will do with their half a million dollars, but they agree that I will end up having half a million dollars from one of them. Still, we wouldn't want to conclude that Peter killed the victim or that I will inherit half a million dollars. Why is this? The conflict involved seems to be harmful not only to the conflicting part itself but harmful to the whole situation as such. The existence of the conflict puts us in doubt about the credibility of the witnesses and makes us suspect that something more general 'has gone wrong.' We might suspect that the two witnesses or the siblings have arranged their statements, or that the conditions for seeing

[10] Horty [4, p. 69] already suggests something similar in his considerations of **Economy** and **Nixon**: "Perhaps the extreme predictions are best seen as undermining each other and the truth lies somewhere in between."

Peter kill the victim weren't that great or that our parents have no intention to reveal anything about who gets their money. This explanation can be made for the **Economy** example, too.[11] In other cases, like **Student Housing** or **Ice-Skating**, the conflict doesn't seem to destroy or harm anything over and above the conflicting part itself. We have some information that speaks in favor of Brigt Rykkje being Norwegian and we have some other information that speaks in favor of Brigt Rykkje being Dutch. However, the different kinds and sources of information are independent of each other and are not destroyed by the conflict. In all of the cases where the conflict is harmful to the general argument, this is so because the *credibility* of the sources of information or the *authority* of the experts is undermined by the conflict.[12] It is not clear, however, how this explanation succeeds to explain the different intuitions about **Mary in Canada** and **Carol in the US**, or **Wavelength Color** and **Primary Color**. The conflict involved is exactly of the same form, and thus, it is not clear why the conflict is harmful for one floating conclusion but not for the other.

4 A possible solution: A default framework for floating conclusions

In the last section, we found that none of the presented hypotheses is suited to explain the intuitions about *all* examples. Still, we are positive that the two hypotheses "Possible Compromise" and "Harmfulness of the Conflict" combined manage to describe what is at the heart of the matter for the different examples. For example, "Possible Compromise" nicely explains the different judgments for **Mary in Canada** and **Carol in the US** and for **Wavelength Color** and **Primary Color** by referring to the compromising proposition.

We propose a solution that manages to combine different hypotheses. The basic idea is that a floating conclusion should be accepted by default. We should prima facie believe them. Then, there are different reasons to deviate from the default and to reject a floating conclusion. One such reason is explained by the "Possible Compromise" thesis. If there is a compromise between the conflicting propositions (and from this compromising proposition the floating conclusion doesn't follow) then the floating conclusion is to be rejected.[13] Another reason to deviate from the default and to reject the floating conclusion is described

[11] It would be interesting to see if the situation changes when the two conflicting propositions that seem to 'cancel each other out' are not equally strong.

[12] This can be seen even clearer when modeling examples like **Yacht** or **Murderer** in a different way. For example, one can take e.g. "Witness A says that p, hence p" to be not a defeasible argument but rather a justification through testimony or utterance for p. The argument as such then starts from the two premises "Peter killed the victim with a knife" and "Peter killed the victim with a gun" (which are both justified by some testimony). Then one could argue that both arguments (from K to V and from G to V) are in fact even *undermined* (see [8] for this terminology) since the premises of the arguments are attacked by the conflicting testimonies. Many thanks to Stipe Pandzic for this remark.

[13] It can be noted that the vagueness thesis describes a special case of a possible compromise.

in the "Harmfulness of the conflict" thesis. If a conflict is not only harmful to the conflicting propositions but undermines the credibility or authority of the sources of information entirely, then the floating conclusion is to be rejected. The basic idea behind this can already be found in [5, p. 189]: "we might suppose that, although floating conclusions are in general acceptable, there are structural features present in situations such as the yacht example, but not yet captured in our formal representations of these examples that block these conclusions."

In all cases of floating conclusions, the conflicting propositions are contrary to each other. Although they cannot both be true at the same time, they can both be false at the same time. That is, in addition to the possibility that one proposition is true and the other false (or vice versa), there is a third possibility: both propositions are false. The two reasons to deviate from the default describe both one version of (or reasons for) this third possibility. Either we reject both propositions because the credibility of their justification has been undermined or because there is a third proposition as a compromise available. [14] Logically, one could capture this by saying that both conflicting propositions p and q that lead to a floating conclusion are false, i.e., $\neg p \wedge \neg q$ or $\neg(p \vee q)$. The other (default) situation in which we should accept the floating conclusion could then be captured by the *exclusive disjunction* $p \veebar q$ of the two propositions being true. [15] With this manifold solution, we think that we manage best to precisely describe what is going on in the different examples and hit the heart of the matter, describing the underlying patterns of the intuitions. In the case of **Primary Color**, **Nixon**, or **Carol in the US** it is really the plausibility of the compromise between the conflicting propositions (either that the cup is orange, that Nixon is politically in the middle, or that Carol lives 'in between') that makes us reject the floating conclusion intuitively. In the cases of **Yacht** and **Murderer** or **Economy** [16] the reason why we intuitively reject the floating conclusion is that we do not trust any line of argument anymore as the credibility of the sources got destroyed. For example, in the **Murderer** case, the credibility of the testimonies is destroyed by their disagreeing about the weapon. Moreover, we do not want to claim that these two reasons: compromise and harmfulness of the conflict are the only reasons for deviating from the default of accepting the floating conclusion. Plausibly, there will be other reasons. This is not a problem for our theory, though, as this default-based framework can easily be extended with multiple more reasons to deviate.

[14] In this sense, the "hidden default" thesis can also be incorporated into the framework. The two explanations "possible compromise" and "harmfulness of the conflict" describe different reasons why in some examples a default still seems to be missing or hidden.

[15] Thanks to an anonymous reviewer for pointing this out.

[16] In fact, **Economy** can be explained both by referring to a possible compromise *and* by the harmfulness of the conflict for the credibility of the sources. Thus, this example shows that there can be even more than one reason to deviate from the default of accepting floating conclusions.

5 Conclusion and Outlook

In this paper, we investigated the phenomenon of floating conclusions. The question about the acceptability of floating conclusions can be reformulated as the question of whether we should accept at least one line of reasoning among a set of conflicting lines of reasoning. We presented an overview of different examples of floating conclusions from the literature and extended the list with new examples. We examined different hypotheses that aim to explain our non-uniform intuitions about whether floating conclusions should be accepted or not and tested them via our examples. We argued that no hypothesis succeeds in explaining our intuitions concerning *all* the presented examples. Instead, we presented an overarching explanation for the acceptability of floating conclusions. The explanation starts with the basic idea that floating conclusions ought to be accepted *by default*. The framework then allows several reasons to deviate from the default and to reject the floating conclusion. These reasons come into play when there seems to be a third alternative besides the two conflicting propositions. We presented two possible reasons why this alternative arises. If there is a compromise between the conflicting propositions from which the floating conclusion does not follow or if the conflict is harmful to the sources of information, one can deviate from the default and reject the floating conclusion. We saw that these two reasons nicely cover and explain all the examples investigated in this paper. We thereby manage to describe the underlying pattern of our intuitions regarding the floating conclusions. Still, the framework is open for new, additional reasons that will come along with new examples when the matter is investigated more.

As we already mentioned in the introduction, intuitions alone do not always help us decide about the different examples. This is visualized nicely in the following example from practical reasoning. Imagine there was a robbery where jewelry was stolen. Later, the police stop a man in a car and find the stolen jewelry. The police have reason to believe that the occupant stole the jewelry. However, the man claims to have bought the jewelry. Both activities (stealing and the so-called 'Hehlerei'/'heling,' i.e., the purchase of stolen goods) are punishable in the Netherlands as well as in Germany. The German legal system allows the suspect to be convicted for the crime with the lesser penalty since it is clear that he committed one of the two crimes. The Dutch legal system, on the other hand, cannot convict the suspect unless there is evidence that clearly shows which of the crimes was committed. [17] The acceptance of the practical floating conclusion (the suspect is punishable) here does not depend on intuitions but on the legal system, one is referring to. The dependency on context and on stakes can also be nicely visualized by our presented examples. While the conflict destroys the credibility of the witnesses in **Yacht** or **Murderer**, the conflict does not seem to destroy the credibility of Anna and Ben in the color examples. In these contexts, where they are simply telling us the color of a cup, we have no reason to be suspicious because the context offers

[17] According to: Hans Nijboer, personal communication, 2007

us no reason why they should lie about the color of the cup. [18] Since whether or not we want to choose the third alternative, deviate from the default, and reject the floating conclusion seems to depend heavily on the stakes and on the context, we consider it a very difficult challenge to represent the appropriate handling of floating conclusions in a formal logical system. Moreover, these examples suggest that there might be a difference between purely theoretical, epistemological reasoning, and practical reasoning. As the intuitions would also become more comparable when actions are involved, further research on the influence of practical reasoning for floating conclusions seems very fruitful.

References

[1] Brandom, R., "Articulating Reasons: An introduction to inferentialism," Harvard University Press, Cambrdige, Massachusetts, 2000.

[2] Broersen, J., *Rethinking the boid with an eye on making it more responsible*, manuscript (unpublished).

[3] Ginsberg, M. L., "Essentials of Artificial Intelligence," Morgan Kaufmann Publishers In, San Francisco, 1993.

[4] Horty, J. F., *Skepticism and floating conclusions*, Artificial Intelligence **135** (2002), pp. 55–72.

[5] Horty, J. F., "Reasons as Defaults," Oxford University Press, Oxford, 2011.

[6] Horty, J. F., R. H. Thomason and D. S. Touretzky, *A skeptical theory of inheritance in nonmonotonic semantic networks*, Artificial intelligence **42** (1990), pp. 311–348.

[7] Makinson, D. and K. Schlechta, *Floating conclusions and zombie paths: two deep difficulties in the "directly skeptical" approach to defeasible inheritance nets*, Artificial intelligence **48** (1991), pp. 199–209.

[8] Pollock, J. L., "Cognitive carpentry: A blueprint for how to build a person," MIT Press, Cambridge, Mass, 1995.

[9] Prakken, H., *Intuitions and the modelling of defeasible reasoning: some case studies*, in: *Proceedings of the 9th International Workshop on Non-Monotonic Reasoning. Toulouse, France*, 2002.

[10] Reiter, R., *A logic for default reasoning*, Artificial intelligence **13** (1980), pp. 81–132.

[11] Sellars, W., *Inference and meaning*, Mind **62** (1953), pp. 313–338.

[12] Veltman, F., "Logics for conditionals," Thesis (1985).

[18] Thanks to Joris Graff for coming up with this point.