Straw man as misuse of rephrase

JACKY VISSER
Centre for Argument Technology / University of Dundee, United Kingdom
j.visser@dundee.ac.uk

MARCIN KOSZOWY
Faculty of Law / University of Białystok, Poland & Centre for Argument Technology / Polish Academy of Sciences, Poland
koszowy@uwb.edu.pl

BARBARA KONAT
Centre for Argument Technology / Polish Academy of Sciences, Poland
bkonat@gmail.com

KATARZYNA BUDZYNSKA
Centre for Argument Technology / Polish Academy of Sciences, Poland & University of Dundee, United Kingdom
k.budzynska@dundee.ac.uk

CHRIS REED
Centre for Argument Technology / University of Dundee, United Kingdom & Polish Academy of Sciences, Poland
c.a.reed@dundee.ac.uk

The 'rephrase' relation between propositions is introduced in Inference Anchoring Theory to facilitate argument mining (the automated analysis of argumentative discourse). Examining an example from the candidates' debates leading up to the 2016 presidential elections in the United States, we explore the relation between such rephrases and the *straw man* fallacy. Our aim with the structural characterisation of the fallacy is to work towards a foundation for the automated identification of

rephrase and inference patterns as a tool in computationally identifying instances of straw man and similar fallacies (e.g. *ignoratio elenchi*).

KEYWORDS: fallacies, Inference Anchoring Theory, rephrase, straw man, television debate, US elections

1. INTRODUCTION

The fallacies have traditionally been a central subject in the field of argumentation studies. Despite continuing progress on the development of computational methods in this field, little work has been done on the computational evaluation of argumentation (Walton, 2016) – that is, the evaluative task, not in terms of persuasive effectiveness or logical validity, but rather in terms of the violation of some conventional (often dialectical) norms for reasonable argumentative conduct. To explore a possible route towards the automated indication of potential fallacies, we propose to characterise some fallacies in structural terms as a pattern of various argumentative relations between propositions. In particular, we focus on the misuse of the rephrase relation by means of which speakers can reformulate their claims and arguments without affecting the inferential structure. The idea is that these patterns are amenable to machine recognition in annotated text corpora or as part of systems for argument mining, and as such can advance the computational evaluation task.

Of course, we do not intend to claim that all instances of the pattern point to fallacies: there are legitimate reasons for, e.g., restating one's argument or claim. However, if these patterns can be identified automatically, then a normative criterion could be brought to bear on these cases to determine their argumentative legitimacy – like other argumentative moves, rephrases can be thought of as operating on a continuum ranging from reasonable to fallacious counterparts (van Eemeren, 2010). In the present paper, we do not concern ourselves with drawing the boundary between reasonable and fallacious, rather we focus primarily on the structural characterisation of a fallacy that we consider to involve rephrase: straw man.

In the next section, 2, we first turn to the notion of 'argumentative rephrase'; how it relates to similar concepts, and its role in our theoretical

framework of Inference Anchoring Theory (IAT). Section 3 covers some relevant existing perspectives on the straw man fallacy, and our structural characterisation of straw man in terms of IAT. In Section 4, we illustrate how this structure is realised in argumentative practice, by discussing an example from a television debate in the lead-up to the 2016 US presidential elections. Before turning to the Conclusion, we describe two directions for future work in more detail in Section 5: extension of the approach to the *ignoratio elenchi* fallacy, and linguistic cues for the calling-out of rephrase-related fallacies.

2. REPHRASE IN ARGUMENTATIVE DISCOURSE

2.1 The backgrounds of argumentative rephrase

Restating one's own words can be a powerful argumentative device – just to name a few benefits: it can help the audience to memorise or to understand what the main claim is; it can give the impression that the speaker has more arguments that he or she in fact has; it can protect the speaker's position from being attacked by starting with a strong claim and then gradually withdrawing to "safe ground" by weakening the standpoint by means of a rephrase.

The notion of 'rephrase' bears some resemblance to other communicative phenomena – e.g. paraphrase, textual entailment and summarisation – but is distinct from those in several important respects. Rephrase is intimately connected with the intentions of the speaker: it could be possible for example, that a nuanced or prolix point is rephrased aggressively, succinctly or discourteously, with a very great distance between the two sentences in lexical and semantic terms – yet both points being clearly recognisable as related through a rephrase. On the other hand, just because one span of discourse (even in a single interaction) entails or paraphrases another, this does not mean that the speaker is, in fact, intentionally using the one to rephrase the other. Thus whilst the surface lexicalisation from which paraphrase and entailment have previously been recognised is indeed also useful in recognising rephrase, more is required.

The *text mining* community has long recognised the importance of paraphrases, i.e. two units expressing similar meaning (see e.g. the Microsoft Research Paraphrase corpus (Dolan, Brockett, & Quirk, 2005)). Hirst (Hirst, 2003) defines paraphrase as "talking about the same situation in a different way" with changes in the wording or syntactic

structure. Bhagat and Hovy (Bhagat & Hovy, 2013) propose the term "quasi paraphrases" to describe text units "that convey approximately the same meaning using different words". In some cases, the existence of paraphrase can be used as an indicator of the existence of a rephrase relation – but it is neither a sufficient nor a necessary condition: some sentence pairs that happen to be paraphrases might not be used as rephrases, and some rephrases might be far to distant lexically and semantically to count as paraphrase.

Textual entailment (Dagan, Glickman, & Magnini, 2006; Berant, Dagan, & Goldberger, 2012; Zanzotto, Pazienza, & Pennacchiotti, 2005; Dagan & Magnini, 2013; Mirkin, Berant, Dagan, & Shnarch, 2010) is also closely related to rephrase inasmuch as some examples of rephrase that are not paraphrase might involve textual entailment relations. Once again, though, the very fact that textual entailments are (according to the leading approaches in that area) definable on the basis of solely the lexicalisation of the two involved text spans, suggests that it is functioning differently to rephrase, where the contextual embedding of the spans provides essential information about whether or not they constitute a rephrase.

2.2 Rephrase in Inference Anchoring Theory

Our characterisation of straw man as an infelicitous use of the rephrase relation is based on Inference Anchoring Theory (IAT) (Budzynska & Reed, 2011). Building on insights from discourse analysis and argumentation studies, IAT explains argumentative conduct in terms of the anchoring of argumentative reasoning in dialogical interaction, by means of the 'illocutionary connection' between the two. Elaborating on traditional speech act theory (Austin, 1962; Searle, 1969), illocutionary forces are reinterpreted as relations connecting locutions to propositional content. The reasoning appealed to in the argumentation involves three types of relations between propositions. An inference relation holds between a proposition that functions as a premise in an argument and the contested proposition that it supports. A conflict relation indicates that a proposition is incompatible with another. A rephrase relation holds between two propositions when one proposition is used to rephrase, restate or reformulate another proposition. Whether there is a rephrase relation between the propositional content of two text spans depends on the speaker's intention to modify the wording of a

premise or a conclusion, by means of, e.g., specialisation, generalisation or instantiation.

Characteristic of IAT is its orientation towards computational linguistic methods and software implementation. To facilitate the required machine-readability, IAT adheres to the extended Argument Interchange Format (AIF$^+$) standard (Chesñevar, McGinnis, Modgil, Rahwan, Reed, Simari, South, Vreeswijk, & Willmott, 2006; Reed, Wells, Rowe, & Devereux, 2008). AIF$^+$ is a graph-based ontology that facilitates the representation of the intertwined locutionary, illocutionary, and propositional structures, resulting from the analysis of argumentative discourse.

The computational orientation comes to the fore in the availability of software to support the use of IAT. For example, OVA+ (Janier, Lawrence, & Reed, 2014) is an online tool for argument analysis facilitating the representation of the structure of argumentative discourse. OVA+ is freely at the website `ova.arg.tech` and has been used to produce all figures in this paper. Analyses produced with OVA+ (and a variety of other programs) can be saved as 'argument maps' in AIFdb (Lawrence, Bex, Reed, & Snaith, 2012), an online searchable repository of analysed arguments freely available at `aifdb.org`. The argument maps stored in AIFdb can subsequently be collected in corpora at `corpora.aifdb.org` (Lawrence & Reed, 2014).

Previous work on the introduction of the rephrase relation into IAT started with an intuitive concept of rephrase used in mediation sessions (Janier & Reed, 2017), leading to a preliminary theoretical model of rephrase (Konat, Budzynska, & Saint-Dizier, 2016). Although more complex constellations are possible, in the simplest case of rephrase two text spans serve (almost) the same function in the argumentation, as is the case in Example (1) from the first television debate for the 2016 US presidential elections (Peters & Woolley, 2016b).

(1) a. WALLACE: *[...] You support a national right to carry law. Why, sir?*

b. TRUMP: *[...] In Chicago, which has the toughest gun laws in the United States, [...] they have more gun violence than any other city. So we have the toughest laws, and you have tremendous gun violence.*

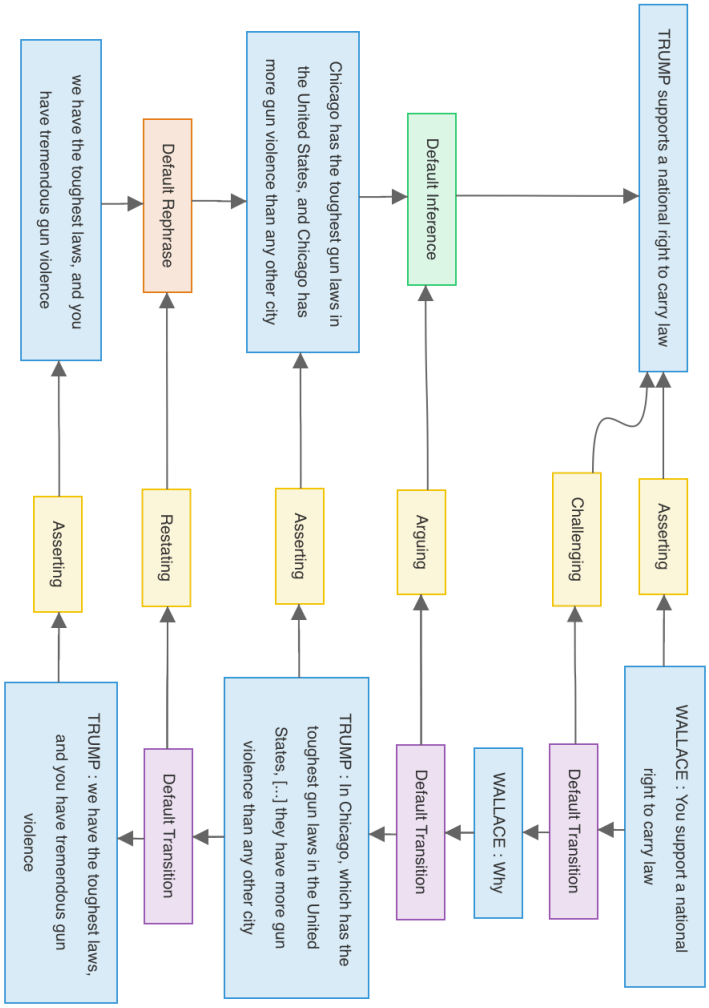

Figure 1 – Diagramming argumentative discourse with Inference Anchoring Theory

The rephrase relation in (then candidate for the Republican party) Donald Trump's response to a challenge from the debate's moderator Chris Wallace constitutes an integral part of the argument's structure but without introducing a new line of argument, as presented in Figure 1. The right-hand side of 1 represents the dialogical dimension of the argumentation, the left-hand side shows the propositional dimension, and the middle row contains the illocutionary connections (see

Section 4.1 for more on this style of annotation with IAT). The right side of Figure 1 shows how Example (1) is segmented into four locutions, interconnected with discourse transitions. On the left of 1, three propositional contents are reconstructed (on the basis of the respective illocutionary connections). The top two propositions function as respectively a conclusion (or claim) and a premise (or argument). The bottom proposition is a rephrase of Trump's argument, not constituting a new line of argument, but rather a restatement or reformulation of the original point.

3. REPHRASE IN STRAW MAN FALLACIES

The aim of this section is to answer the question: how the rephrase relation is structurally related to the straw man fallacy? The answer will be given by employing the general method of modelling rephrase in IAT, as discussed in section 2.2 to capture instances of *Default Rephrase* and *Default Conflict* in the straw man technique. This task would require determining how distinctive structural features of straw man can be depicted using IAT structures represented with OVA+.

3.1 The straw man fallacy

In this section we will expose those key features of the straw man fallacy that serve as a source of inspiration for modelling it as misuse of rephrase. Although straw man is a dialogical strategy aimed at pursuing different goals (Aikin & Casey, 2011; Macagno & Walton, 2017) and thus there is a variety of types of straw man, the common way of defining this technique is to associate it with the misrepresentation of someone's position in order to easily refute that position (Walton, 1996; Talisse & Aikin, 2006; Lewiński, 2011; Lewiński & Oswald, 2013; Aikin & Casey, 2016). This general idea of misrepresentation is related to rephrasing because a rephrase might constitute means to modifying original speaker's standpoint in such a way that it is made easier to attack. According to Macagno and Walton, dialogical purposes of employing the straw man technique consist of "attacks to the argument or the claim of the interlocutor in a dialogue (real or fictitious) in order to reject it, and possibly thus supporting the opposing one" (Macagno & Walton, 2017, p. 110). Diagrammatic representation of straw man proposed in this paper allows us to show a functional role of rephrase relation in fallacious attacks on other party's claim or argument.

As Oswald and Lewiński point out, the often overlooked feature of the straw man technique is its meta-discursive role, meaning that "it operates on someone else's discourse that serves as a material for linguistic maneuvering'' (Oswald & Lewiński, 2014). Putting an emphasis on this meta-discursive feature of straw man leads to conceiving this technique as "a certain unreasonable, sometimes manipulative, re-interpretation of another position''. Our approach relies on emphasising this observation in the graphical representations of instances of straw man fallacy. For the meta-discursive function of straw man is shown in OVA+ diagrams by introducing the rephrase relation in terms of the difference between the propositional content of original speaker's locution and the rephrased statement that has been attacked. This way of representing the meta-discursive feature of straw-man will be shown in Section 3.2.

3.2 The IAT-rephrase approach to straw man

The rephrase approach to the straw man technique can be described in following steps: firstly, the proponent claims C, next, the opponent restates the claim as C' making it easier to attack and finally, the opponent attacks C'. When focusing on dialogical aspects of straw man fallacy, we characterise it as a sequence of two dialogue moves, the first of which is an instance of restating the other party's claims, and the second an instance of attacking the rephrased content. This characteristics allows us to expose the dynamics of the dialogue involving instances of straw man.

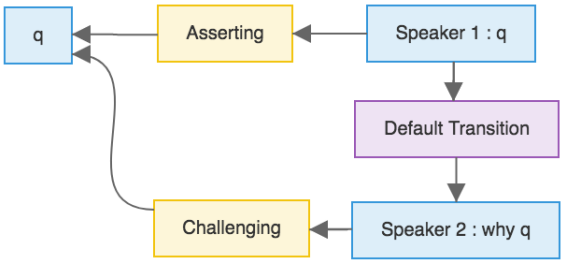

Figure 2 – Non-fallacious challenging

Figure 2 represents an instance of non-fallacious challenging, leading to a non-mixed difference of opinion, where a standpoint is met with a position of doubt. Of course, there is also potential for a straw man

to be committed in in explicit disagreement. In this case, instead of a mere expression of doubt, a stronger position is taken by expressing a contrary standpoint. A non-fallacious version of this pattern is shown in Figure 3

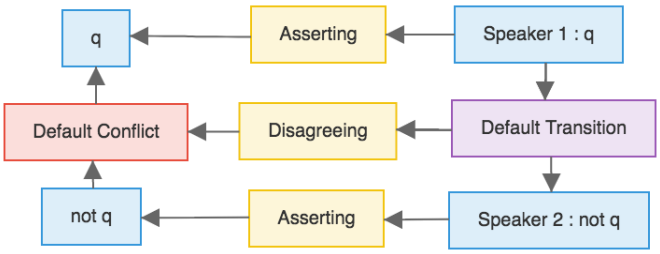

Figure 3 – Non-fallacious disagreeing

Whereas an interlocutor can reasonably challenge or disagree with the propositional content of an original locution, a requirement for this is that the propositional content that is challenged or disagreed with is represented fairly. This is where the rephrase relation comes in. Outwith formal dialogue systems, it would be too strong to expect ordinary language users to express their doubt or disagreement with respect to an exact repetition of the proposition as it was expressed by their interlocutor. Rather, we might expect that the propositional content is restated to some degree. Our suggestion is that there is a continuum of rephrase ranging from literal repetition of a proposition to expressing an entirely different proposition, with the extremes not entailing rephrase and everything in between being an instance of rephrase to some degree. Now, different approaches to rephrase (and its related notions, such as paraphrase and fuzzy quantifiers in logic) will define this continuum in different ways. With respect to the reasonableness of rephrases in the challenging and supporting of standpoints, and their potential to lead to the straw man fallacy, it requires a normative theory of argumentation to determine where on this continuum the boundary may be between reasonable and fallacious rephrase use. Independent of where the boundary of reasonableness is drawn, the structural constellation of interacting propositions, illocutionary connections and locutions would instantiate a common pattern. In Figures 4 and 5 we show two such patterns.

The first pattern, as shown in Figure 4, represents a potential straw man in challenging. It differs from the pattern of non-fallacious challenging (as represented in Figure 2) with respect to the *Default*

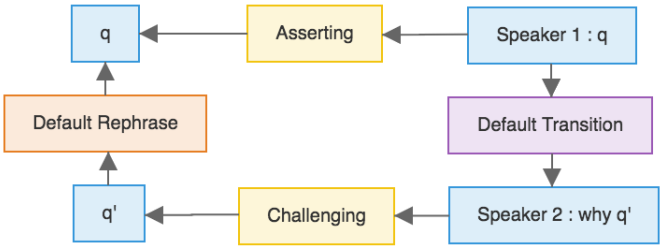

Figure 4 – Potential straw man in challenging

Rephrase node on the left hand side of the diagram and the fact that the illocutionary connection of *Challenging* does not target the content q of the original speaker's locution, but instead the rephrased content q'. These two elements may constitute the structural cue indicating that the straw man technique might have been employed.

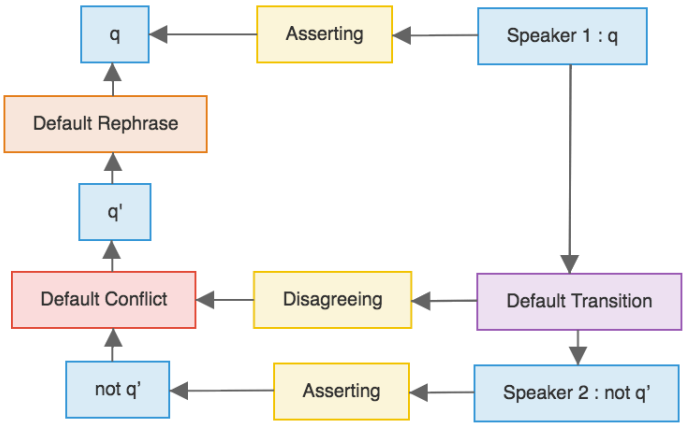

Figure 5 – Potential straw man in disagreeing

The second pattern represented by Figure 5 shows how the straw man technique may be structurally combined with disagreeing. This pattern differs from non-fallacious disagreeing, as shown in Figure 3 with respect to the fact that q has been first rephrased as q' and next that the illocutionary connection of *Disagreeing* targets the *Default Conflict* relation between q' and $not - q'$ on the left hand side of the diagram. In case of non-fallacious disagreeing there is no *Default Rephrase* relation and the *Default Conflict* relation is between q and $not - q$.

4. EMPLOYING REPHRASE STRUCTURES TO IDENTIFY POTENTIAL FALLACIES IN PRACTICE: THE 2016 US PRESIDENTIAL ELECTIONS

4.1 Annotation of the 2016 US presidential election television debates

To explore how the proposed structural characterisation of straw man can be applied in practice, we will consider an example from the television debates leading up to the 2016 US presidential elections. A selection of three of the televised election debates have been fully annotated on the basis of IAT. The annotated transcripts of the first general election debate, and of the two preceding first television debates for the primaries of the Republicans and of the Democrats, are collected in the US2016tv corpus (which is available online at `corpora.aifdb.org/US2016tv`).[1] The transcripts together amount to 58,900 words, which are segmented into a total of 4,671 locutions. The annotation contains 4,277 propositions, connected through 1,551 inference relations, 194 conflict relations, and 333 rephrase relations.

The US2016tv corpus is annotated by four annotators. The annotators are extensively trained to analyse the television debates on the basis of IAT (see Section 2.2), resulting in an inter-annotator agreement on a 10.5% (word count) sample of a Combined Argument Similarity Score κ (Duthie, Lawrence, Budzynska, & Reed, 2016) of 0.679 (the usual Cohen's κ (Cohen, 1960) is less suitable for this kind of complex multi-layered annotation task, yielding a value of 0.516). Below, we summarise the annotation guidelines – the full version of which deals with, among others: anaphoric references, epistemic modalities, repetitions, punctuation, discourse indicators, interposed text, reported speech, and how to deal with context-specific peculiarities.

Segmentation divides the (transcribed) text into locutions. A locution consists of a speaker designation and an 'argumentative discourse unit' – a text span with discrete argumentative function (Peldszus & Stede, 2013).

Transitions capture the functional relationships between locutions, reflecting the dialogue protocol – a high level specification of the set of transition types that are available in a particular

---

[1]While all of the examples we present in this paper are drawn from the 2016 presidential debates, some come from debates that are not included in the US2016tv corpus. They are, however, all annotated in accordance to the guidelines we summarise here. The two larger examples, (2) and (3), are available in a separate corpus at `corpora.aifdb.org/FallaciousRephrase`.

communicative activity.

Illocutionary connections embody the intended communicative functions of locutions or transitions, such as: *Agreeing*, *Arguing*, *Asserting*, *Challenging*, *Disagreeing*, *Questioning*, *Restating*, and *Default Illocuting* (when none of the other types suffice). Some types of illocutionary connection lead to the reconstruction of a propositional content. Illocutionary connections can anchor in locutions or in transitions, depending on their type and on possible indexicality (*viz*. the different anchoring of *Challenging* in Figures 1 and 2).

Inferences are directed relations between propositions, reflecting that a proposition is meant to supply a reason for accepting another proposition. An argument scheme (e.g., *Argument from Example* or *Argument from Expert Opinion*) can be specified, failing that, it is labelled as *Default Inference*.

Conflicts are directed relations between propositions, reflecting that a proposition is meant to be incompatible with another proposition or relation. Such incompatibility may depend on, e.g., logical contradiction or pragmatic contrariness, or the annotated relation may default to *Default Conflict*.

Rephrases are directed relations between propositions, reflecting that a proposition is meant to be a reformulation of another proposition. Such reformulation may involve, e.g., *Specialisation*, *Generalisation* or *Instantiation*, or the relation defaults to *Default Rephrase*.

4.2 The radical Islam example

Example (2) illustrates the relation between rephrase structures and straw man (Peters & Woolley, 2015). In this example, the debate's moderator, John Dickerson, asks (then candidate for the Democrat presidential nomination) Hilary Clinton to reflect on what Marco Rubio (then candidate for the Republican nomination) has said at an earlier occasion about being at war with radical Islam. In her response, Clinton first restates Rubio's claim that has been cited by Dickerson and next she disagrees with this rephrased content. These dialogue moves are visualised diagrammatically in Figure 6 with the use of the OVA+ tool mentioned in Section 2.2.

(2) a. DICKERSON: *Secretary Clinton, you mentioned radical jihadists. Marco Rubio, also running for president, said that this attack showed [...] that we are at war with radical Islam.*

Do you agree with that characterization, radical Islam?

b. CLINTON: *I don't think we're at war with Islam. I don't think we're at war with all Muslims. [...]*

c. DICKERSON: *Just to interrupt. He didn't say all Muslims. He just said radical Islam.*

Dickerson, in his move (2-a) reports on what Rubio has said about radical Islam. Reported speech is represented on the top of the diagram presented in Figure 6 with two locutions and one propositional content. Within the same move, Dickerson asks Clinton to express her opinion on Rubio's claim that has just been reported. In Clinton's response (2-b), the original phrase "radical Islam" has been replaced first with "Islam" and then with "all Muslims". Clinton disagrees here with Rubio by saying "I don't think we're at war with Islam". Figure 6 represents this move with the illocutionary force of *Disagreeing* which is anchored in the transition between the two locutions. Then Clinton restates her own statement by saying "I don't think we're at war with all Muslims", which is represented in Figure 6 with a *Default Rephrase* node on the left-hand side of the diagram. The illocutionary force of *Restating* links together the transition between the two locutions with *Default Rephrase*.

Employing OVA+ to the analysis of this example allows us to pinpoint the structural overlap between the argumentative function of straw man and the dialogical structure of rephrase. Clinton, by saying "I don't think we're at war with Islam" (see utterance (2-b)) introduces the rephrased propositional content to the dialogue, namely "this attack showed that we are at war with Islam" (see the second top propositional content on the left hand side of the diagram). As this is the content assigned to Rubio that Clinton disagrees with, in the diagram there is a *Disagreement* node that is linked to the *Default Conflict* node between two contradictory statements, namely: "this attack showed that we are at war with radical Islam" and "we are not at war with Islam". Structurally speaking, if the straw man has not been committed, there would be a *Default Conflict* relation between Rubio's claim "this attack showed that we are at war with Islam" and its negation, e.g. "we are not at war with Islam". This analysis also shows that the dialogical complexity of the straw man technique may consist of the use of the two different instances of the rephrase relation. The second instance of rephrasing is represented in Figure 6 at the bottom left of the diagram. The statement "we are not at war with Islam" has been rephrased again as "we are not at

war with all Muslims". The use of two rephrasing dialogue moves related to one instance of the straw man fallacy shows how this technique may be further developed by rephrasing the content that is already an instance of rephrasing.

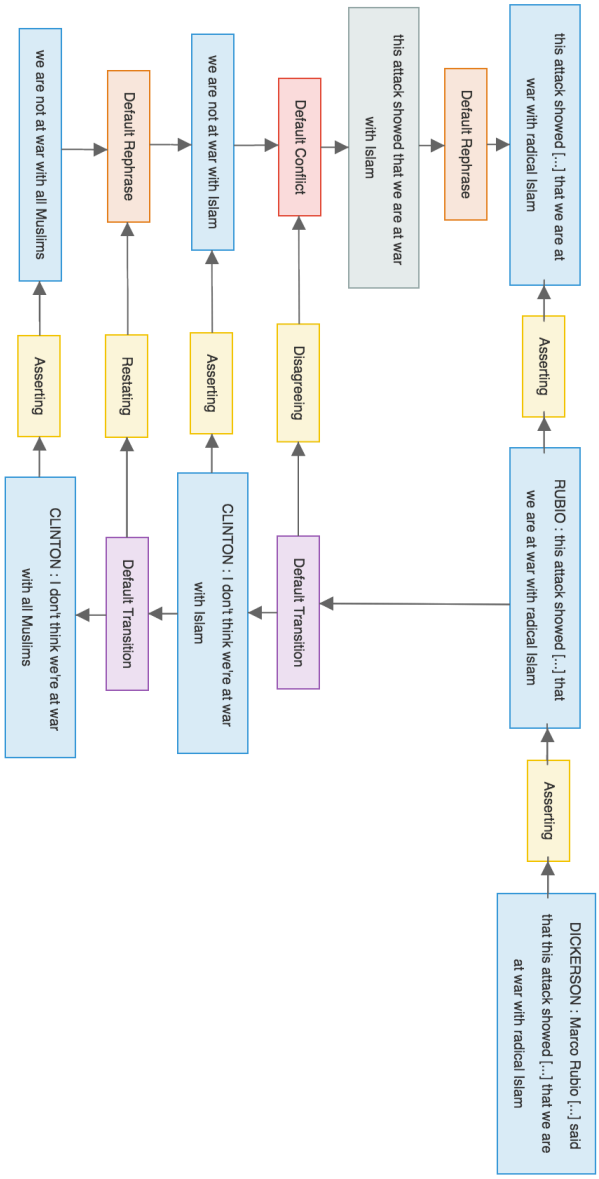

Figure 6 – Diagrammatic visualisation of Example (2)

To sum up, the OVA+ diagram helps to structurally pinpoint the straw man technique as an illocutionary force of disagreeing with the rephrased propositional content instead of disagreeing with the original content. In terms of the conflicts between the propositional contents, Figure 6 helps to identify an instance of a straw man by showing that the *Default Conflict* relation instead of targeting the original propositional content ("we are not at war with radical Islam") targets the content that has been rephrased ("we are not at war with Islam"). This structural difference allows us to emphasise the meta-discursive feature of a straw-man technique that has been discussed in Section 3.1. As this main structural feature of a straw man fallacy is unique in terms of OVA+ structures, the proposed method of structurally representing instances of a straw man fallacy may be a point of departure for an automatic extraction of these particular structures from large repositories of natural language texts.

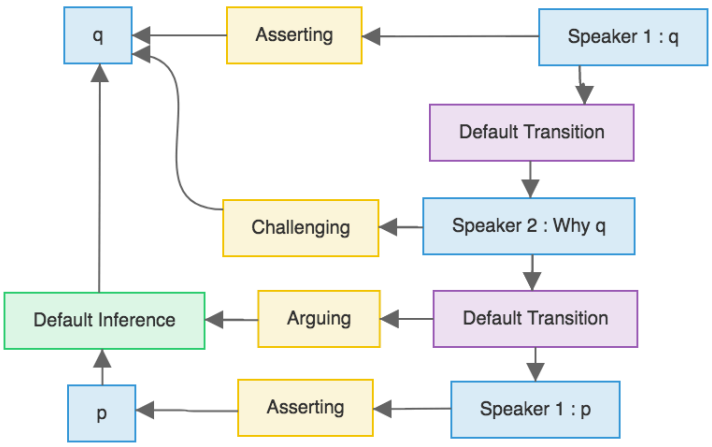

Figure 7 – Non-fallacious supporting

5. PERSPECTIVES ON FUTURE WORK

5.1 Extension to ignoratio elenchi

The rephrase approach to straw man can be extended to other fallacies that rely upon a misuse of the rephrase relation. As a case in point, we can consider *ignoratio elenchi*. While our (narrow) interpretation of straw man involves the restating of an interlocutor's claims, ignoratio elenchi

can be narrowly interpreted as the fallacious rephrase of the speaker's own claim. Ignoratio elenchi is a fallacy with various (and rather varying) interpretations. We propose to presently consider a version of ignoratio elenchi that can be conceived of as the mirror image of the straw man fallacy as we understand it in the current paper: a rephrase of one's own standpoint in order to make it easier to defend (*viz*. the straw man fallacy involving the rephrase of an opponent's standpoint to make it easier to attack). A proponent commits an ignoratio elenchi when making a claim C, and then argumentatively defending C', where C' is a rephrase of C making it easier to defend.

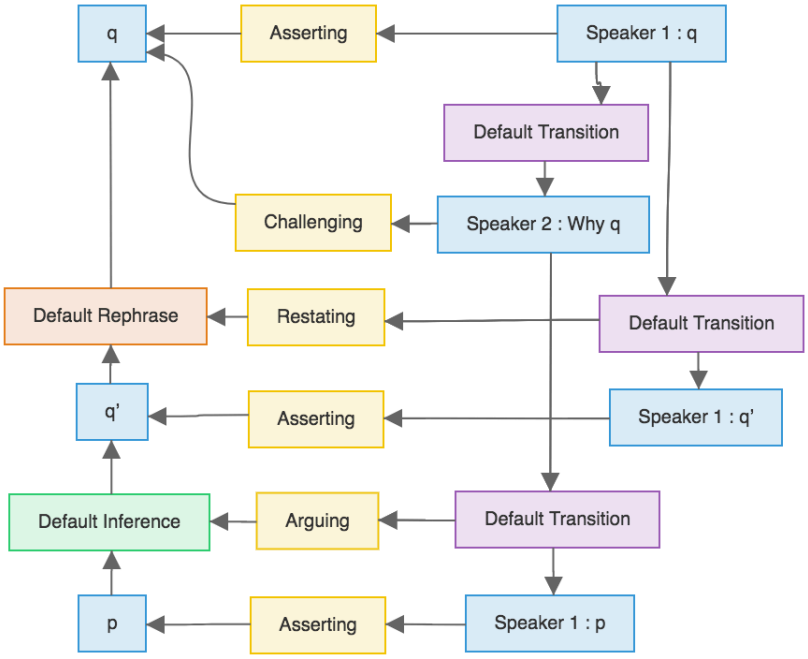

Figure 8 – Potential ignoratio elenchi in supporting

In contrast to the diagrammatic visualisation of a reasonable dialogical counterpart of ignoratio elenchi in Figure 7, Figure 8 shows how in our conception of ignoratio elenchi a proposition q' is introduced standing in a rephrase relation to the original propositional content of the standpoint q. This rephrase relation between q' and q is anchored via an illocutionary connection of *restating*, but this is not part of the ongoing dialogue transitions. The argument p is a dialogical continuation of the

request for argumentative defence (i.e. the challenge of q), while the rephrase as q' is on a separate dialogical track; a dead end as it is unconnected to the provision of p as an argumentative defence instigated by the challenge.

In Example (3), we can see this pattern of interacting rephrase and inference relations occurring, as visualised in Figure 9 (Peters & Woolley, 2016a). During the first head to head debate between presidential candidates Trump and Clinton, the moderator Lester Holt, challenges Trump on his earlier claim that Clinton does not look presidential. Trump quickly restates this into a rephrased claim about Clinton not having the required stamina, and arguing why this is of such importance for a president.

(3) a. HOLT: *Mr. Trump, this year Secretary Clinton became the first woman nominated for president by a major party. Earlier this month, you said she doesn't have, quote, "a presidential look." [...] What did you mean by that?*
b. TRUMP: *She doesn't have the look. She doesn't have the stamina. I said she doesn't have the stamina. And I don't believe she does have the stamina. To be president of this country, you need tremendous stamina.*
c. HOLT: *The quote was, "I just don't think she has the presidential look."*
d. TRUMP: *[...] You have so many different things you have to be able to do, and I don't believe that Hillary has the stamina.*
e. [...]
f. CLINTON: *You know, he tried to switch from looks to stamina.*

5.2 Calling out rephrase fallacies

Both examples discussed in sections 4.2 and 5.1 contain dialogue moves that do not belong to the rephrasing structure, but which may serve as discourse cues indicating that rephrase-related fallacies might have been committed. The common feature of these dialogue moves is what we label 'calling out rephrase fallacies' which consists of attempts at signalling that a misuse of rephrase has just occurred in one of the previous dialogue moves.

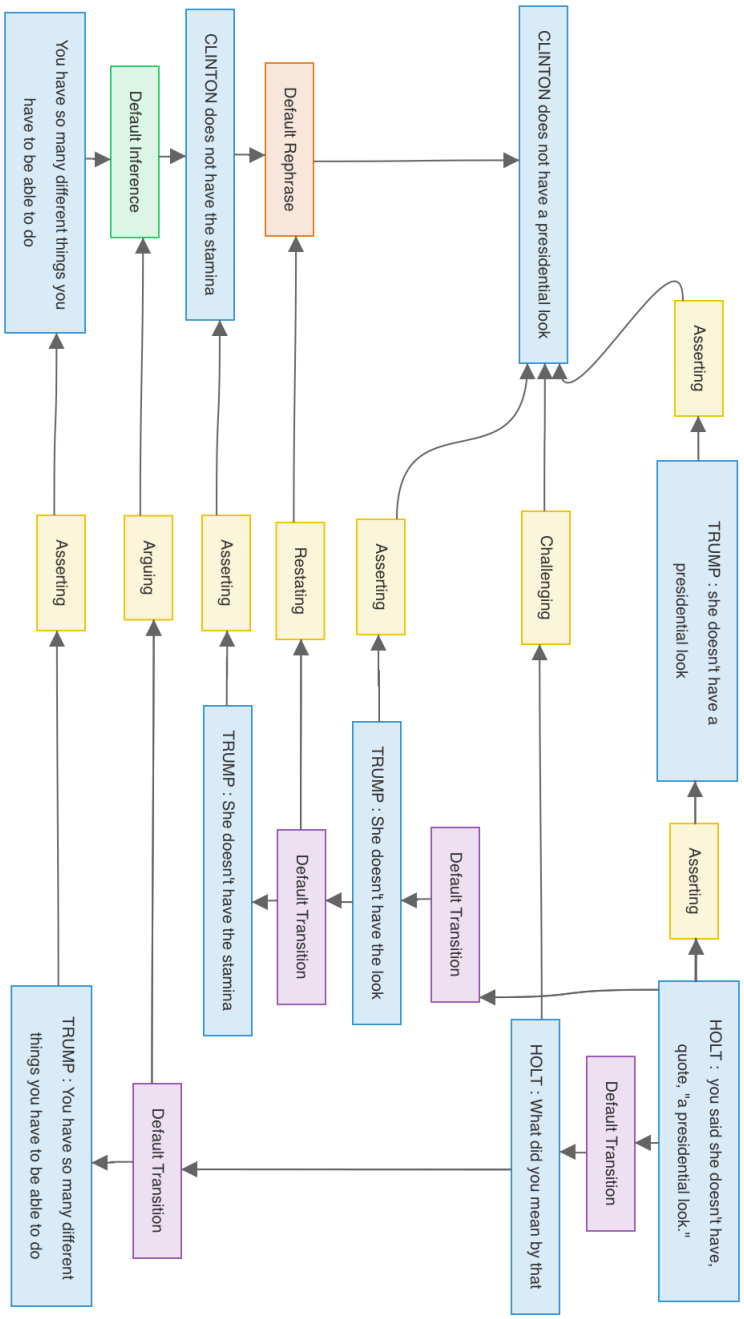

Figure 9 – Diagrammatic visualisation of Example (3)

In example (2) discussed in Section 4.2, Dickerson, by saying "He didn't say all Muslims. He just said radical Islam" (see dialogue move (2-c)), is putting an emphasis on the fact that there was a switch from Rubio's reported claim containing the phrase "radical Islam" to Clinton's view about the rephrased "all Muslims" (see dialogue move (2-c)). Although this content does not strictly belong to the rephrase structure, it shows what type of dialogue moves can be performed when a speaker's aim is to identify an instance of a straw man fallacy. A similar function has Clinton's dialogue move (3-f) shown in example 9 in Section 5.1. By saying "You know, he tried to switch from looks to stamina", Clinton indicates that an attempt at committing a rephrase-related fallacy has been made. Analysing these and other examples of calling out may serve as a preparatory study aimed at exploring linguistic cues for identifying rephrase-related fallacies. For instance, utterances such as "he didn't say…" or "he tried to switch…" may be helpful in finding instances of potential misuses of rephrase in large text repositories.

6. CONCLUSION

Exploring the foundations of automated fallacy detection, we have presented a structural rephrase-based characterisation of the straw man fallacy. By narrowly conceiving of straw man in terms of a misuse of a rephrase of the proposition being challenged, the resulting patterns can facilitate automated techniques for their recognition. Using Inference Anchoring Theory as the framework for our characterisation awards our present theoretical work a close connection to computational implementation. By means of examples from the 2016 presidential election debates in the US we have shown that the patterns we present for the straw man and ignoratio elenchi fallacies do indeed occur in annotated corpora of argumentative discourse. While our structural characterisation does not provide a normative criterion for evaluating any occurrences as reasonable or fallacious, normative theories of argumentation could tell that the examples we discussed can be evaluated as fallacious.

Our exploratory work gives rise to new questions that can be addressed in further research (in addition to that described in Section 5). Some of these questions are of a theoretical or analytical nature, such as: Where are the boundaries of reasonable and fallacious rephrase? How can implicitness of argumentative moves camouflage straw man and

other rephrase fallacies? Does rephrase systematically play a part in other fallacies, e.g. circular reasoning? Other questions relate to the computational aspects we lightly touched upon: How can implicitness in natural language be dealt with in argument mining? How can the automated search for structural patterns be implemented? How can we devise computational methods to distinguish between fallacious and reasonable argumentation?

Beyond leading to such questions of an academic nature, working towards the automated recognition of rephrase fallacies, such as straw man, can turn out to have huge societal impact. Amidst ongoing worries about the prevalence of 'fake news', more and more emphasis is put on fact-checking in journalism and the media at large. While the current focus of the fact-checkers is squarely on determining the truthfulness of a specific bit of information, it appears to us that the way that information is phrased is at least of equal importance. This is especially the case since it often involves the checking of a politician's or other public figure's words, and a negative outcome of the fact-checking effectively amounts to an ethotic attack. It goes without saying that it is of the utmost importance in this endeavour that the actual words spoken are not (intentionally or unintentionally) rephrased to such an extend that it becomes a straw man, and neither should a speaker be allowed to weasel their way out of a negative fact-checking outcome by committing an ignoratio elenchi.

ACKNOWLEDGEMENTS: This research was supported in part by the Engineering and Physical Sciences Research Council in the United Kingdom under grant EP/N014871/1, and in part by the Polish National Science Centre under grant 2015/18/M/HS1/00620. We would like to thank Mark Snaith for his technical help in preparing this paper in LaTeX.

REFERENCES

Aikin, S., & Casey, J. (2011). Straw men, weak men, and hollow men. *Argumentation*, *25(1)*, 87–105.

Aikin, S., & Casey, J. (2016). Straw men, iron men, and argumentative virtue. *Topoi*, *35*, 431–440.

Austin, J. L. (1962). *How to Do Things with Words.* Clarendon Press.

Berant, J., Dagan, I., & Goldberger, J. (2012). Learning entailment relations by global graph structure optimization. *Computational Linguistics*, *38*(1), 73–111.

Bhagat, R., & Hovy, E. (2013). What is a paraphrase?. *Computational Linguistics*, *39*(3), 463–472.

Budzynska, K., & Reed, C. (2011). Whence inference. Tech. rep., University of Dundee.

Chesñevar, C., McGinnis, J., Modgil, S., Rahwan, I., Reed, C., Simari, G., South, M., Vreeswijk, G., & Willmott, S. (2006). Towards an argument interchange format. *The Knowledge Engineering Review*, *21*(04), 293–316.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37–46.

Dagan, I., Glickman, O., & Magnini, B. (2006). The PASCAL recognising textual entailment challenge. In *Machine learning challenges. evaluating predictive uncertainty, visual object classification, and recognising tectual entailment*, pp. 177–190. Springer.

Dagan, I., & Magnini, B. (2013). Entailment graphs for text exploration. In *Proceedings of the Joint Symposium on Semantic Processing. Textual Inference and Structures in Corpora*.

Dolan, B., Brockett, C., & Quirk, C. (2005). Microsoft research paraphrase corpus. *Retrieved March 2005*.

Duthie, R., Lawrence, J., Budzynska, K., & Reed, C. (2016). The CASS Technique for Evaluating the Performance of Argument Mining. *Proceedings of the 3rd Workshop on Argument Mining, Association for Computational Linguistics.*, 40–49.

Hirst, G. (2003). Paraphrasing paraphrased. In *Invited talk at the ACL International Workshop on Paraphrase, Sapporo*.

Janier, M., Lawrence, J., & Reed, C. (2014). OVA+: An argument analysis interface. In Parsons, S., Oren, N., Reed, C., & Cerutti, F. (Eds.), *Proceedings of the Fifth International Conference on Computational Models of Argument (COMMA 2014)*, pp. 463–464, Pitlochry. IOS Press.

Janier, M., & Reed, C. (2017). Towards a theory of close analysis for dispute mediation discourse. *Argumentation*, *31*(1), 45–82.

Konat, B., Budzynska, K., & Saint-Dizier, P. (2016). Rephrase in argument structure. In *Foundations of the Language of Argumentation COMMA 2016 Workshop*. ACM.

Lawrence, J., Bex, F., Reed, C., & Snaith, M. (2012). AIFdb: Infrastructure for the Argument Web. In *Proceedings of the Fourth International Conference on Computational Models of Argument (COMMA 2012)*, pp. 515–516.

Lawrence, J., & Reed, C. (2014). AIFdb Corpora. In Parsons, S., Oren, N., Reed, C., & Cerutti, F. (Eds.), *Computational Models of Argument*, Frontiers in artificial intelligence and applications, pp. 465–466, Netherlands. IOS Press.

Lewiński, M. (2011). Towards a critique-friendly approach to the straw man fallacy evaluation. *Argumentation*, *25(4)*, 469–497.

Lewiński, M., & Oswald, S. (2013). When and how do we deal with straw men? a normative and cognitive pragmatic account. *Journal of Pragmatics*, *59(B)*, 164–177.

Macagno, F., & Walton, D. (2017). *Interpreting Straw Man Argumentation: The Pragmatics of Quotation and Reporting*. Springer International Publishing.

Mirkin, S., Berant, J., Dagan, I., & Shnarch, E. (2010). Recognising entailment within discourse. In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, pp. 770–778. Coling 2010 Organizing Committee.

Oswald, S., & Lewiński, M. (2014). Pragmatics, cognitive heuristics and the straw man fallacy. In Herman, T., & Oswald, S. (Eds.), *Rhétorique et cognition: perspectives théoriques et stratégies persuasives*, pp. 313–343. Peter Lang, Bern.

Peldszus, A., & Stede, M. (2013). From argument diagrams to argumentation mining in texts: a survey. *International Journal of Cognitive Informatics and Natural Intelligence (IJCINI)*, *7*(1), 1–31.

Peters, G., & Woolley, J. T. (2015). Democratic candidates debate in Des Moines, Iowa, November 14, 2015.. Accessed 11 Aug. 2017.

Peters, G., & Woolley, J. T. (2016a). Presidential debate at Hofstra University in Hempstead, New York, September 26, 2016.. Accessed 11 Aug. 2017.

Peters, G., & Woolley, J. T. (2016b). Presidential debate at the University of Nevada in Las Vegas, Nevada, October 19, 2016.. Accessed 11 Aug. 2017.

Reed, C., Wells, S., Rowe, G., & Devereux, J. (2008). AIF+: Dialogue in the argument interchange format. In Besnard, P., Doutre, S., & Hunter, A. (Eds.), *Proceedings of the 2nd International Conference on Computational Models of Argument (COMMA 2008)*, pp. 311–323. IOS Press.

Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language.* Cambridge University Press.

Talisse, R., & Aikin, S. (2006). Two forms of the straw man. *Argumentation*, *20(3)*, 345–352.

van Eemeren, F. (2010). *Strategic Maneuvering in Argumentative Discourse: Extending the Pragma-dialectical Theory of Argumentation*. John Benjamins.

Walton, D. (1996). The straw man fallacy. In van Benthem, J., van Eemeren, F., Grootendorst, R., & Veltman, F. (Eds.), *Logic and Argumentation*, pp. 115–128. Royal Netherlands Academy of Arts and Sciences, Amsterdam.

Walton, D. (2016). Some artificial intelligence tools for argument evaluation: An introduction. *Argumentation*, *30*(3), 317–340.

Zanzotto, M. F., Pazienza, T. M., & Pennacchiotti, M. (2005). Discovering entailment relations using "textual entailment patterns". In *Proceedings of the ACL Workshop on Empirical Modeling of Semantic Equivalence and Entailment*, pp. 37–42. Association for Computational Linguistics.