

Reference: Schumann J., Zufferey S. & Oswald S. (2019): What makes a straw man acceptable? Three experiments assessing linguistic factors. *Journal of Pragmatics* 141, 1-15. <https://doi.org/10.1016/j.pragma.2018.12.009>

What makes a straw man acceptable? Three experiments assessing linguistic factors

Jennifer Schumann¹, Sandrine Zufferey² & Steve Oswald³

A straw man is a form of fallacious argument that involves the distortion of an opponent's view in order to make it more extreme and therefore less acceptable, thus easier to attack. In three experiments, we assess linguistic factors that may influence the acceptability of straw men for addressees. In Experiment 1, we demonstrate that a straw man is more acceptable when the speaker's argument is attacked rather than his standpoint. In Experiment 2, we show that a straw man is more acceptable when it contains two arguments that are simply juxtaposed rather than linked by a causal connective indicating attributed content (i.e. the French *puisque*). In Experiment 3, we find that a straw man is more acceptable when it echoes the speaker's explicit rather than implicit meaning. Taken together, these experiments strongly indicate that the formulation of a straw man has an impact on its perceived acceptability. Our results thus open a new avenue of inquiry for studies assessing the role of linguistic factors on the acceptability of fallacies.

Keywords: argumentation; fallacies; straw man; discourse connectives; French; experimental studies

¹ jennifer.schumann@rom.unibe.ch
University of Bern, Institut de Langue et de Littérature françaises, Länggassstrasse 49, CH – 3012 Bern, Switzerland.

² sandrine.zufferey@rom.unibe.ch
University of Bern, Institut de Langue et de Littérature françaises, Länggassstrasse 49, CH – 3012 Bern, Switzerland.

³ steve.oswald@unifr.ch
University of Fribourg, English Department, Av. de l'Europe 20, CH – 1700, Switzerland.

1. Introduction

When people are involved in argumentative discussions, they make use of different types of arguments to convince, some of them more legitimate than others. Arguments that do not follow normative standards of argumentative discourse are referred to as fallacious arguments (e.g. Hamblin, 1970). Among these, a specific case is that of the straw man fallacy, which typically involves a “misrepresentation of someone’s position in order to easily refute that position” (Oswald and Lewiński, 2014). Let us imagine a situation in which four friends, Andrea, Bryce, Carol and David, are discussing some plans for the weekend. Andrea and Bryce are in a heated discussion about leisure activities, while Carol and David are calmly chatting a few steps apart. Andrea informs Bryce that she thinks his proposal of going skydiving is not really a good idea. Bryce swiftly turns to the others claiming “Hey guys, looks like we’re going to spend a very quiet weekend, since Andrea doesn’t want us to have fun!” This example illustrates the core feature of a straw man fallacy: the misrepresentation of the linguistic content of another’s utterance. Bryce, in repeating Andrea’s statement, has distorted her skepticism about the leisure activity into the more extreme position of refusing to engage in any fun activity. Thus, the straw man fallacy is characterized by a dichotomy of form and function (see Lewiński, 2011). This misrepresentation is the form under which the straw man operates in order to attack the opponent, Andrea, and refute her stance on the topic. The straw man lies in the distortion of the original content, which can occur under various forms (see Aikin & Casey, 2011, 2016) and it aims at attacking the opponent with a more extreme, and therefore weakened, version of the original statement. In recent years, the straw man fallacy has attracted considerable attention in argumentation theory (see below). However, contrary to other cases of fallacies, the characteristics of the straw man that make it more or less acceptable as an argument have so far not been experimentally tested. The aim of this paper is to provide an empirical evaluation of the role of three linguistic elements involved in straw man fallacies as factors increasing or decreasing their acceptability.

The paper is structured as follows. In Section 2, we present studies that have assessed the straw man fallacy from various theoretical perspectives. In Section 3, we discuss the open research questions that arise from this state-of-the-art and that will form the basis of our experimental studies. Section 4 reports the first experiment, designed to

test whether the locus of the misrepresentation included in the straw man, namely the opponent's standpoint or argument, has an influence on its acceptability, which is related to participants' ability to detect the fallacy. The second study, presented in section 5, concentrates on the role of discourse connectives that indicate attributed content, such as the French causal connective *puisque* (a connective resembling the English *since* or *given that*) as a guide to help hearers detect straw man fallacies. In experiment 3, reported in Section 6, we assess the distinction between a misrepresentation bearing on the opponent's explicit or implicit meaning in order to assess whether surface form similarities increase hearers' perceived familiarity with the fallacious argument and thus its acceptability. We summarize the global picture emerging from our experiments and discuss new research perspectives for experimental studies on the straw man fallacy in section 7.

2. Previous studies on the straw man fallacy

Fallacies represent a well-established topic in argumentation theory (see Hamblin, 1970; van Eemeren & Grootendorst, 1992; Hansen & Pinto, 1995; Walton, 1995; Tindale, 2007). Even though the straw man fallacy has been theorized by pragma-dialectics early on (van Eemeren & Grootendorst, 1992), the number of studies devoted to this fallacy has strongly increased only in recent years (see Aikin & Casey, 2011, 2016; Lewiński, 2011; Lewiński & Oswald, 2013; Macagno & Walton, 2017; Oswald & Lewiński 2014; Saussure, 2018).

Among contemporary models, the standard pragma-dialectical model of argumentation, put forward by van Eemeren & Grootendorst (1992, 2004), offers a normative framework in which fallacies may be precisely characterized. Through the theorization of argumentation as a critical discussion, pragma-dialecticians define a series of 10 rules whose observance guarantees the reasonableness of argumentation conducted in view of resolving a difference of opinion. In this approach, fallacies are accordingly described as violations of these rules for critical discussion. The straw man fallacy, in particular, amounts to a violation of rule 3, the standpoint rule, which states that speakers must attack the standpoint that was actually put forward by their opponent: when speakers attack a distortion of the original standpoint, they commit the straw man fallacy (van Eemeren & Grootendorst, 1992, Chapter 11). Interestingly,

even if it does make room for considering straw man attacks on arguments,⁴ this classical treatment of the fallacy formulates rule 3 only in relation to standpoints. Our first experiment will provide evidence favoring the view that straw man fallacies are actually less detected, and therefore more likely to succeed in fooling their addressees, when they target an argument than when they target a standpoint (see sections 4.5 and 4.6 below).

Over the years, and in order to accommodate the observation that arguers still try to win arguments while trying to uphold standards of critical reasonableness, pragma-dialecticians have updated their framework into an integrated pragma-dialectical model with the notion of *strategic maneuvering* (van Eemeren & Houtlosser, 1999; van Eemeren, Grootendorst & Snoeck Henkemans, 2002; van Eemeren et al., 2014). This notion refers to the maneuvering possibilities that arguers have at their disposal within the normative space delimited by the operation of the rules for critical discussion: they are indeed able to shape and formulate their argumentative moves in a way that is rhetorically optimal for their purpose, without this necessarily resulting in the fallaciousness of said moves. However, the moment such maneuvering violates any of the 10 rules, it is said to derail, which leads to the production of a fallacy. Typically, there are three ways in which speakers may strategically maneuver: they may linguistically and discursively shape their message by trying to (i) meet *audience demand* (tailor their argumentative moves according to what they believe the audience will be more likely to agree with), (ii) control the *topical potential* so as to define and delimit the topics under discussion to their advantage, and (iii) select the *presentational devices* more likely to fulfill their rhetorical goals (this level of strategic maneuvering is typically focused on linguistic choices). This evolution of the model makes it relevant to account for the straw man fallacy in the way we approach it experimentally here, because it focuses on how rhetorical opportunities may be linguistically shaped.

Both the pragma-dialectical treatment of the straw man fallacy and the accounts inspired by it – but that extend it (like Lewiński's) – identify, rightly in our view, the crucial feature of the fallacy, namely misrepresentation. By discussing how arguers may attribute distorted or even fictitious standpoints to their opponents, and by assessing how misrepresentations may happen (exaggeration, omission of nuances

⁴ See van Eemeren & Grootendorst (1992, p. 43) and Lewiński (2011, p. 479) for a discussion.

and qualifications, simplification, absolutization, misuse of quantification, etc.), these models rightly emphasize the truly pragmatic nature of this particular fallacy and foreground the importance of linguistic choices in the persuasive enterprise.

The research we report on here may thus be seen as an experimental contribution to this effort, as we aim to identify specific linguistic constructions and devices that may positively impact the success of the straw man fallacy. Yet, the added value of our experimental investigation of the straw man fallacy is to be found in a cognitively-grounded account of the mechanisms of misrepresentation that seem to be at its core.⁵ Closer to our interest in cognitive processes, another fruitful approach to the study of fallacies has come from approaches which make use of the framework of relevance theory, and in particular those which integrate the notion of epistemic vigilance (Sperber et al., 2010), or who combine pragma-dialectics and relevance theory (Lewiński & Oswald, 2013; Oswald & Lewiński, 2014). These different frameworks discuss a series of factors that may make a straw man effective. According to Lewiński & Oswald (2013) the straw man is effective when an opponent is not aware of the fallacy. In order to remain unnoticed, the interpretation therefore needs to be limited to plausible contextual assumptions that appear relevant and leave out critical information. While Lewiński and Oswald (2013) characterize general strengthening and weakening strategies meant to increase the straw man's undetectability, they offer neither typological descriptions of linguistic expressions typically used in straw man fallacies, nor experimental evidence for these. Saussure (2018) considers the straw man fallacy to be a prestige-gaining device that has the ability to remain a winning move even if the fallacy fails as a persuading move. In this view, the straw man is effective as an *ethos* boosting device, because it leads addressees to draw inferences about the speaker's skillfulness as a conversationalist and about the target's weakness, who ends up bearing the burden of proof with no real possibility to defend her/himself. While these hypotheses can clearly be related to our own project, notably in terms of the compatibility of our respective frameworks, their scope on speaker's *ethos* is not at the core of the present investigation.

⁵ To be fair, pragma-dialecticians are very-well aware of this and acknowledge that for the type of externalized analysis they intend to carry out, cognitive insights are not needed (see van Eemeren & Grootendorst, 2004, p. 74).

What emerges, therefore, is that none of these approaches has discussed in depth the linguistic structure and necessary verbal components of a straw man fallacy. Determining the typical linguistic characteristics of the straw man fallacy that could lend themselves to empirical testing is therefore a task that remains to be carried out. This being said, we would like to state that we do take inspiration from research in the framework of relevance theory (Sperber & Wilson, 1995) to identify which linguistic components of the straw man might influence its acceptability in the next section.

Empirical approaches to the study of fallacies have flourished over the past decade, with studies focusing for example on the abusive *ad hominem* fallacy (van Eemeren, Garssen & Meuffels, 2012), the slippery slope (Lillo-Unglaube, Canales-Johnson, Navarrete & Fuentes Bravo, 2014), or the *ad populum* fallacy (Ozols, Maillat & Oswald, 2016). However, to our knowledge, only one empirical study by Bizer, Kozak and Holterman (2009) has tested the persuasiveness of the straw man fallacy. Their first experiment has shown that personal relevance (i.e., how meaningful the content was for the participants) has an impact on the straw man's effectiveness. In a context of low personal relevance, the straw man turned out to be more persuasive than in a context of high relevance. In the high relevance case, the participants paid more attention because the argumentative material appeared to be more important to them. The results of the second experiment have shown that the personal need for cognitive closure does also influence the straw man's effectiveness in a context of high relevance. Participants with a high need for structure tended to quickly define their position and be more prone to be influenced by the straw man. It appears from these studies that the straw man fallacy does not seem to be universally persuasive. A striking point is that even though the straw man fallacy is a meta-discursive fallacy, hence a fallacy that operates by misrepresenting the discourse of an opponent, neither of the experiments focuses on the role of linguistic factors in the persuasive force of the straw man. This begs the question: apart from individual differences and personal relevance, are there also linguistic factors that increase or decrease a straw man's effectiveness? That is, are there linguistic ways to make misrepresentation pass for genuine, faithful, representation? In our experiments, we tackle this question by first determining some specific linguistic factors that may have an influence on its acceptability, as we will describe in the next section, and then by putting them to empirical testing in the rest of this paper.

3. Operationalization of linguistic factors and hypotheses

In order to get a more fine-grained understanding of how the straw man fallacy works, we identified three linguistic factors we think are typically involved in straw man fallacies and tested their impact on the straw man's acceptability. Let us however first note, for the sake of conceptual and terminological clarity, that we take acceptability and undetectability to be closely related in the processing of fallacies, the idea being, following Hamblin's classical definition of fallacies (1970), that failure to spot a fallacious argument is likely to increase the chances that the argument is accepted.

The first factor is the role of discourse connectives. In the framework of relevance theory, connectives are described as procedural markers guiding the hearer's interpretation of discourse (e.g. Blakemore, 1987, 2002). As such, they increase textual coherence and help hearers to better understand and remember the logical links between discourse segments (Caron, Micko & Thüring, 1988; Millis, Golding & Barker, 1995; Sanders & Noordman, 2000). By using causal connectives in argumentative discourse, speakers can linguistically indicate the causal link between a standpoint and an argument – be it sound or fallacious, as in the case of the straw man fallacy. The interesting feature of causal connectives from an argumentative perspective is that various causal connectives introduce different types of causal relations between discourse segments (e.g. Pit, 2007; Zufferey, 2012). For example, in French, the difference between the causal connective *parce que* (the closest translation equivalent for the English *because*) introduces a cause presented as new information in the discourse context. By contrast, the connective *puisque* (closest to the English is *since* and *given that*) acts as an accessibility marker (Zufferey, 2014) that indicates that the cause following the connective contains given information, in other words information that is already known or accessible in the hearer's cognitive environment. A similar point was made by van Eemeren, Houtlosser and Snoeck Henkemans (2007) in their pragma-dialectic framework about the connective *since*, which also has the ability to implicitly convey that the starting point for the discussion, hence the information following the connective, has mutually been accepted by the speaker and the opponent. In our experiments, we make use of this feature of *puisque* in order to check whether hearers use its procedural information to detect that the argument following the connective is thus being tacitly attributed to another speaker. If

hearers use this information, we hypothesize that this should alert them to the possibility that the attributed content is not accurate, and therefore help them to detect the straw man, compared to a situation in which the attributed content is introduced without an explicit connective.

Another crucial characteristic of the connective *puisque* is that it is used to introduce a justification rather than a cause (Franken, 1996), thus increasing the hearer's perception that the speaker is trying to justify their position. When the segment introduced by the connective contains a straw man, these expectations are not de facto met because the fallacious argument represents new information that does neither echo nor justify the already stated content. This should raise hearers' awareness and lead to a better detection of the fallacy. In a nutshell, we hypothesize that the straw man fallacy will appear to be more acceptable when the attributed content is juxtaposed rather than introduced by the connective *puisque*, because the connective announces attributed content and should stimulate the participants' epistemic vigilance filters⁶.

The second factor that we tested in our experiments concerns the way a fallacious argument can be formulated. More specifically, the straw man can either make use of another speaker's explicit meaning – and in that case the wording of the fallacious argument remains as close as possible to the original utterance – or rely on another speaker's implicit meaning. In the latter case, the wording differs substantially from the original argument and the link between them must be recovered by inference. It should be stressed that the distinction that we make between implicit and explicit meaning is solely based on the wording of the straw man, more specifically how faithful the target argument is with respect to the original argument. The closer a reformulation is to the utterance it echoes, the more explicit it is. This definition should not be mixed up with the distinction between explicitly and implicitly conveyed content, put forward by Oswald and Lewiński (2014) regarding the straw man fallacy, as the latter use the explicit/implicit distinction to denote the type of content in the original utterance that is more likely to be misrepresented without this being noticed – Oswald & Lewiński's

⁶ In recent years, research on epistemic vigilance (Sperber et al., 2010) has claimed that humans possess a suite of cognitive mechanisms meant to evaluate other people's claims and to assess verbal messages in terms of both their content and their source.

assumption being that addressees are less likely to spot misrepresentations of implicatures than misrepresentations of explicatures, for commitment attribution reasons (see Oswald & Lewiński, 2014; Morency et al., 2008). Recent research in the domain of discourse processing led us to expect that straw man fallacies that provide renditions of another speaker's explicit meaning should be more acceptable than those that rely on implicit meaning. In particular, this research has demonstrated that discourse processing is usually a shallow process (Sanford, 2002; Sanford & Graesser, 2006; Sanford & Sturt, 2002; Sanford, Sanford, Molle & Emmott, 2006) as in most cases all words are not interpreted fully, sometimes giving rise to the phenomenon of semantic illusions (e.g. Erickson & Mattson, 1981). An utterance can therefore appear for a hearer to be understandable even if every word is not fully processed, leading to an incomplete representation.

The effectiveness of semantic illusions can be affected by different factors such as communicative relevance (Sanford, 2002) working memory load (Bohan, 2008), information structuring (Sanford & Sturt, 2002; Sanford, Price & Sanford, 2009), and word similarity (Wang, Hagoort & Yang, 2009). Using the factor of word similarity, we hypothesized that straw man fallacies would be less detected when the misrepresentation was based on the speaker's explicit rather than implicit meaning, because the former creates an illusion of familiarity, thereby potentially increasing the chances of acceptance. In order to test this factor, in one of our experiments, the fallacious content was conveyed either in an implicit or an explicit way. The implicit straw man contained an inaccurate reformulation of the original utterance, so that the original content was misrepresented. The explicit straw man fallacy was built on a wording that remained as close as possible to the original statement. This was achieved by replacing only one lexical element of the subordinate clause containing the argument, in order to create an illusion of semantic similarity that could lead to the fallacy remaining undetected.

As we were designing a controlled format for the cases of straw man fallacies that would be included in our experiments, we noticed that two types of distortions of a speaker's view could be achieved in short dialogues. One targets the opponent's standpoint (i.e., the claim that the speaker puts forward, which is the claim that he wants the addressee to accept) and another one focuses on his argument (i.e., the

information that the speaker provides in order to support the claim). We hypothesized that the locus of the misrepresentation could have an impact on the acceptability of the straw man fallacy, as an attack on the standpoint and an attack on the argument do not have the same argumentative consequences. This line of thinking echoes Pollock's (1987) work on defeasible reasoning. In his account, two types of attacks are recurrent: 1) direct attacks on the conclusion (rebutting defeaters) and 2) attacks on the evidence in support of the conclusion (undercutting defeaters). These two types of attacks bear on what we call respectively the standpoint and the argument. When targeting the standpoint, the attack is of a higher order because it implies a denial of all that the opponent has affirmed, whereas an attack on the argument can leave the standpoint untouched and therefore the discussion can still proceed. Attacking an argument is, indeed, an attack on a reason for a standpoint and not necessarily on the standpoint itself, and in these cases the standpoint may still prevail, albeit for other reasons than the one that was successfully attacked. Conversely, it is possible to accept an argument without necessarily accepting a standpoint. A recent study from Uzelgun, Mohammed, Lewiński & Castro (2015) provides an example. Their research focuses on the use of *yes, but...* constructions in argumentative discourse. They demonstrate that one key characteristic of these constructions is their ability to indicate partial agreement by accepting the content of the argument in the *yes*-clause and refuting the overall standpoint in the *but*-clause. Crucially, the concessions that are made in the *yes*-clause are prone to illegitimate reformulations such as straw man fallacies. We therefore expect that straw man fallacies that target the opponent's arguments should be more acceptable than straw man fallacies that target the opponent's standpoint. We started our experimental series with an experiment testing the distinction between misrepresented arguments and misrepresented standpoints, as we now outline.

4. Experiment 1: misrepresented argument vs. misrepresented standpoint

4.1 Participants

We recruited 30 participants (17 female, mean age: 33, age range 18-60) with higher education (none of them in linguistics) by word-of-mouth. They were all native speakers of French, living in the French speaking part of Switzerland.

4.2 Materials

Participants read 40 short dialogues in French presented in a randomized order. The first part of the dialogue, uttered by a person named Barbara, contained a standpoint and an argument connected by the French causal connective *parce que*, as illustrated in (1).

- (1) Barbara: Il est crucial de mieux soutenir les jeunes parents parce qu'avoir un enfant signifie beaucoup de charges financières.

Barbara: It is crucial to better support young parents because having a child means a lot of financial charges.

This initial sentence of the dialogue remained unchanged in all experimental conditions. It was followed by replies by another person named Alexandre (the names were kept constant throughout the experiment) that came in four different experimental conditions. The second part of the dialogue was designed to test the effect of two different variables. The first variable was the fallacious or non-fallacious nature of the argument presented by Alexandre and the second was the difference between replies that targeted the standpoint and replies that targeted the argument presented by Barbara. The critical segment, where the experimental manipulation took place, was always located in the segment following the connective *puisque* in the response formulated by Alexandre. We now illustrate these four experimental conditions.

The condition involving a misrepresented argument is illustrated in (2). In this condition, the first segment ('let's raise the family allowance') contains a possible consequence that could be derived from Barbara's claim, and the second part of the sentence ('it is only about the money') presents a justification for this consequence in the form of a misrepresented argument. Indeed, Barbara has claimed that having a child implies a certain financial burden but this argument has been transformed to become more extreme (only money matters) thus representing a case of straw man.

- (2) Alexandre: Augmentons les allocations familiales puisqu'on ne pense qu'à l'argent.

Alexandre: Let's raise family allowance since it is only about the money.

By contrast, the second experimental condition, illustrated in (3), implies a non-fallacious reformulation (here, a semantically faithful rendition) of Barbara's initial

argument in the segment following the connective *puisque*. In this condition, the first part of the segment was identical to (2) and the only part that changed between these two conditions was the segment following the connective, in order to control as much as possible the effect of the straw man itself rather than other linguistic elements.

(3) Alexandre: Augmentons les allocations familiales puisque les parents sont sous pression économique.

Alexandre: Let's raise the family allowance since the parents are under economic pressure.

In the third experimental condition, illustrated in (4), the segment that followed the connective contained a misrepresentation of Barbara's standpoint (i.e. 'it is crucial to better support young parents') rather than her argument (i.e. 'having a child means a lot of financial charges') as in (2) and (3). In this condition, the first part of Alexandre's reply contained a possible consequence of Barbara's standpoint and the second segment following the connective contained a misrepresentation of this standpoint, thus again creating a straw man. As in previous conditions, the segments were linked with the connective *puisque*.

(4) Alexandre: Nous devons mieux distribuer l'argent du gouvernement puisque les familles ne savent pas se débrouiller seules.

Alexandre: We have to better distribute the government's money since the families do not know how to cope by themselves.

Finally, the fourth condition is illustrated in (5). It contained a non-fallacious reformulation of Barbara's standpoint. In this condition, the first segment was identical to the one in (4) but this time, the segment following the connective contained a non-fallacious reformulation (we need to help families) of the standpoint defended by Barbara (i.e. 'it is crucial to better support young parents').

(5) Alexandre: Nous devons mieux distribuer l'argent du gouvernement puisqu'il faut aider les familles.

Alexandre: We have to better distribute the government's money since we need to help the families.

The different conditions were assigned to four different lists using a Latin square design, in order to ensure that each item could only appear in one condition. Participants were randomly assigned to one of the lists. All the participants read ten items in all four conditions (40 items in total).

4.3 Procedure

The experiment was set up on the survey platform *SurveyMonkey* and sent to the participants via a web link. The survey started with some preliminary instructions about the experiment. The participants were informed that they would have to read a series of dialogues on political topics and would be asked to respond to four questions for each example. They were also instructed to take their time reading the dialogues and to respond in a spontaneous way to the questions. They then had to answer four demographic questions regarding gender, age, native language and place of residence, before moving on to the experiment itself. There was no time limit set for the completion of the survey – on average the participants took 30 minutes to complete the task.

For each item, the participants had to respond to four questions on a 6-point Likert scale ranging from “No, absolutely not” to “Yes, absolutely”, with an additional option (“I don’t know”) that they could select when they weren’t able or willing to give an answer. The four questions, reproduced in (6) to (9) below, were formulated to assess, in three different and indirect ways, whether the participants had spotted the fallacy or not.

- (6) Est-ce que la conclusion d'Alexandre est proportionnée par rapport à ce qu'a affirmé Barbara ?
Do you think the conclusion given by Alexandre is proportionate to what Barbara has said?
- (7) Est-ce que la conclusion tirée par Alexandre découle logiquement de ce qu'a affirmé Barbara ?
Do you think the conclusion given by Alexandre logically derives from what Barbara has said?
- (8) Êtes-vous d'accord avec Alexandre ?
Do you agree with Alexandre?
- (9) Êtes-vous d'accord avec Barbara ?
Do you agree with Barbara?

Each question tests a different feature that we assume to be linked to the fallaciousness of the straw man. The first question in (6) aims to assess whether the participants had spotted the exaggerative nature of the straw man fallacy. The second question in (7) is meant to assess whether participants realized that there was no logical link between the two statements in the case of straw men. The third question in

(8) targets the participants' agreement with Alexandre. This question is also meant to assess participants' detection of fallacies, as we hypothesize that participants should agree less with Alexandre if they (intuitively) spot that his argument is not legitimate. We expect answers to these three questions to converge, but comparing results obtained on each of them will also allow us to identify whether some aspect of the straw man fallacy is less detectable than the others. By contrast, the fourth question in (9) targets the participants' agreement with Barbara and is designed as a control, as this agreement typically depends on participants' prior beliefs rather than on the argumentative moves presented in the dialogue.

4.4 Analysis

We used a 2x2 repeated measure ANOVA with two within-subject factors: the nature of the argument (normal vs. fallacious) and the element targeted in the reformulation (argument vs. standpoint).

Only the responses given on the scale from 1 to 6 were included in the data analysis. When participants chose the additional option ("I don't know"), their answers were treated as missing data and removed from the analysis.

4.5 Results

In Table 1, we report the mean score of acceptability for each condition, as well as the standard deviation.

Table 1: experiment 1 - misrepresented argument vs. misrepresented standpoint

Question 1: Proportion	Mean	Standard Deviation
Fallacious argument	3.145	0.758
Fallacious standpoint	2.858	0.802
Non-fallacious argument	4.043	0.619
Non-fallacious standpoint	3.889	0.612
Question 2: Logical Link	Mean	Standard Deviation
Fallacious argument	3.389	0.792
Fallacious standpoint	3.097	0.621
Non-fallacious argument	4.028	0.589
Non-fallacious standpoint	3.816	0.609

Question 3: Agreement with Alexander	Mean	Standard Deviation
Fallacious argument	3.321	0.697
Fallacious standpoint	3.053	0.776
Non-fallacious argument	4.212	0.575
Non-fallacious standpoint	3.989	0.539
Question 4: Agreement with Barbara	Mean	Standard Deviation
Fallacious argument	4.334	0.600
Fallacious standpoint	4.153	0.520
Non-fallacious argument	4.360	0.559
Non-fallacious standpoint	4.153	0.477

For the first question, the analysis of variance returned two main effects. First, non-fallacious responses were better accepted than fallacious ones [$F1(1, 29) = 56.640, p < 0.001$; $F2(1, 39) = 139.319, p < 0.001$]. Second, the response was more likely to be accepted when it followed-up on the argument rather than on the standpoint [$F1(1, 29) = 7.991, p < 0.01$; $F2(1, 39) = 5.516, p < 0.05$].

The same two main effects were also found for the second question, targeting the logical nature of the link in the argumentative move. The acceptability was higher for the normal than for the fallacious conditions [$F1(1, 29) = 81.174, p < 0.0001$; $F2(1, 39) = 79.248, p < 0.001$]. Agreement was also higher when the response was a follow-up of the argument than the standpoint [$F1(1, 29) = 11.451, p < 0.01$; $F2(1, 39) = 6.632, p < 0.05$].

For the third question, targeting the agreement with Alexandre, the analysis produced the same pattern. Once again, normal reformulations had a higher acceptance rate compared to fallacious reformulations [$F1(1, 29) = 63.686, p < 0.001$; $F2(1, 39) = 69.011, p < 0.001$] and a higher acceptability rate when the continuation was based on the argument rather than the standpoint [$F1(1, 29) = 10.021, p < 0.01$; $F2(1, 39) = 3.310, p = 0.770$].

No main effect and no interaction were found between fallacious and non-fallacious statements for question 4, targeting the agreement with Barbara [$F1(1, 29) = 0.17, p = 0.899$; $F2(1, 39) = 0.003, p = 0.959$], but in this case as well, participants did prefer statements targeting the argument compared to the standpoint [$F1(1, 29) = 7.021, p = 0.013$; $F2(1, 39) = 5.616, p = 0.023$].

4.6 Discussion

This experiment focused on the difference of acceptability between a misrepresentation of the argument and a misrepresentation of the standpoint. Our results show that the locus of the misrepresented content leads to different detection rates. More specifically, arguments lead to higher acceptance rate compared to standpoints; in other words, the fallacy is less detected when it targets an argument than when it targets the standpoint. In addition, our results also indicate that participants prefer non-fallacious statements, which tends to demonstrate that overall, fallacies are intuitively detected.

The lower detection rate for arguments compared to standpoints is likely due to the fact that the refutation of an argument happens on a lower discursive level than the refutation of a standpoint and does not seal the difference of opinion between speakers. Indeed, when a speaker distorts the argumentative load of the statement made by his opponent, the opponent's standpoint can remain valid – it is only part of the statement that is weakened, but not the statement as a whole. For example, in the experimental item presented above, someone may disagree that parents must be given more money to raise their children (the speaker's argument) but still agree that they need more help of another kind, for instance in the form of an increase in social worker positions meant to help families (the speaker's standpoint).

Our results also indicate that responses to questions 1 to 3 are all consistently dependent on the argumentative move presented in the dialogue. By contrast, responses to question 4 clearly differed from the responses to the previous questions. As expected, in that case it is not the argumentative move that is evaluated, but responses rather reflect the participants' own world view. The contrast between these conditions indicates that participants' responses to questions 1 to 3 represent a valid reflection of their understanding of the argumentative move presented in the dialogue. Based on these results, we decided to only use straw man fallacies that targeted the argument rather than the standpoint in subsequent experiments, as these seem to be more effective.

5. Experiment 2

5.1 Participants

We recruited 33 first-year psychology students (25 female, mean age: 22, age range 18-59) from the University of Fribourg online for this experiment. None of the participants took part in experiment 1. They all resided in the French speaking part of Switzerland and were native speakers of French. The students received 0.25 ECTS for their participation in the experiment.

5.2 Materials

We used the same experimental design as in the previous experiment. The participants saw 40 short dialogues on various political topics in a randomized manner. As shown in (10), we reused the first part of the dialogue from the previous experiment:

- (10) Barbara: Il est crucial de mieux soutenir les jeunes parents parce qu'avoir un enfant signifie beaucoup de charges financières.
Barbara: It is crucial to better support young parents because having a child means a lot of financial charges.

As in experiment 1, the first part of the dialogue, always initiated by Barbara, was identical in all four conditions.

The second part of the dialogue, always uttered by Alexandre, contained the manipulated variable. In this experiment we tested the effect of two factors: the difference between fallacious and non-fallacious statements and the difference between segments containing a reformulation that was introduced by the French connective *puisque* and segments that were juxtaposed without the use of a connective. As in the previous experiment, the first segment ('Let's raise the family allowance') expressed a possible consequence of the argument put forward by Barbara in (10). This part of Alexandre's reply to Barbara remained unchanged in all four conditions. In all experimental items, the straw man fallacy was related to Barbara's argument rather than her standpoint, as this condition was found to yield higher acceptance scores in Experiment 1. We now comment on the four conditions used in this experiment.

The first condition, presented in (11), contained a straw man where the fallacious part was introduced with *puisque*. In order to create a straw man, we distorted Barbara's

initial argument in the second segment ('it only is about the money') resulting in an exaggeration compared to the original one ('having a child means a lot of financial charges').

- (11) Alexandre: Augmentons les allocations familiales puisqu'on ne pense qu'à l'argent.
Alexandre: Let's raise the family allowance since it only is about the money.

In the second condition, illustrated in (12), we removed the connective and presented the same two parts that were simply juxtaposed.

- (12) Alexandre: Augmentons les allocations familiales. On ne pense qu'à l'argent.
Alexandre: Let's raise the family allowance. It only is about the money.

By contrast to the conditions containing straw man as in (11) and (12), we reformulated the original argument put forward by Barbara in a non-fallacious way in (13) and (14). In the third condition, illustrated in (13), we introduced the reformulated argument with the connective *puisque*.

- (13) Alexandre: Augmentons les allocations familiales puisque les parents sont sous pression économique.
Alexandre: Let's raise the family allowance since the parents are under economic pressure.

In the last experimental condition, illustrated in (14), we also removed the connective *puisque* from the non-fallacious version in (13) and presented the two arguments by juxtaposition.

- (14) Alexandre: Augmentons les allocations familiales. Les parents sont sous pression économique.
Alexandre: Let's raise the family allowance. The parents are under economic pressure.

As in Experiment 1, we used a Latin square design to create four different lists that were randomly assigned to each participant. We attributed only one of the four possible conditions to each list, so that every participant saw 10 items per condition (40 items in total).

5.3 Procedure

We used the *Qualtrics*⁷ platform for this experiment and the survey was sent via web link to the participants.

The procedure was identical to the one followed in the previous experiment. The preliminary instructions, as well as the demographic questions, remained the same as in experiment 1. The participants had no time limit to complete the survey and took approximately 30 minutes for the completion of the task.

We asked the participants the same four questions, reproduced in (6) to (9) above, to which they had to respond on a 6-point Likert scale with an additional option ('I don't know').

5.4 Analysis

We used a 2x2 repeated measure ANOVA with two within-subject factors. The first factor targeted, like in Experiment 1, the nature of the argument (normal vs. fallacious) and the second one the way we introduced the reformulated argument (with or without the connective *puisque*).

As in the previous experiment, only the responses given on the scale from 1 to 6 were included in the data analysis. We treated the answers on the additional option ('I don't know') as missing data and removed them from the analysis.

5.5 Results

The mean score of acceptability for all the conditions, as well as the standard deviation are presented in Table 2.

Table 2: experiment 2 – straw man with *puisque* vs. straw man without connective

Question 1: Proportion	Mean	Standard Deviation
Fallacious <i>puisque</i>	3.119	0.597
Fallacious no connective	3.329	0.589
Non-fallacious <i>puisque</i>	4.024	0.561
Non-fallacious no connective	4.163	0.564

⁷ This was the preferred option at the University of Fribourg in Switzerland, where the second and third experiment were conducted.

Question 2: Logical Link	Mean	Standard Deviation
Fallacious <i>puisque</i>	3.524	0.612
Fallacious no connective	3.604	0.533
Non-fallacious <i>puisque</i>	4.048	0.498
Non-fallacious no connective	4.187	0.703
Question 3: Agreement with Alexander	Mean	Standard Deviation
Fallacious <i>puisque</i>	3.128	0.649
Fallacious no connective	3.334	0.613
Non-fallacious <i>puisque</i>	4.043	0.467
Non-fallacious no connective	4.221	0.468
Question 4: Agreement with Barbara	Mean	Standard Deviation
Fallacious <i>puisque</i>	4.510	0.583
Fallacious no connective	4.537	0.535
Non-fallacious <i>puisque</i>	4.717	0.479
Non-fallacious no connective	4.572	0.502

For the first question, the analysis returned two main effects. First, non-fallacious responses were rated as more acceptable than fallacious ones [$F1(1, 32) = 87.526, p < 0.001$; $F2(1, 39) = 44.011, p < 0.001$]. Second, the straw man fallacy was more likely to be accepted when the fallacious segment was not introduced by the connective *puisque* but simply juxtaposed to the previous segment [$F1(1, 32) = 6.736, p < 0.05$; $F2(1, 39) = 7.570, p < 0.01$].

The same pattern was also found for the question targeting the logical link between the statements. The analysis showed that non-fallacious arguments were better accepted than fallacious ones [$F1(1, 32) = 53.527, p < 0.001$; $F2(1, 39) = 24.661, p < 0.001$]. Regarding the second effect, the participants showed some preference for the straw man when the fallacious part was juxtaposed to the previous segment rather than introduced by the connective *puisque*, but for this question the effect was only approaching significance. [$F1(1, 32) = 3.520, p = 0.07$; $F2(1, 39) = 3.228, p = 0.08$]. We report the same tendency for the third question on the agreement with Alexandre. The results confirmed the first effect: the non-fallacious conditions were better accepted than fallacious ones [$F1(1, 32) = 113.451, p < 0.001$; $F2(1, 39) = 37.911, p < 0.001$]. We noticed higher acceptability rates for fallacious arguments introduced without *puisque*. In this case, the effect was significant only in the item analysis [$F1(1, 32) = 3.495, p = 0.07$; $F2(1, 39) = 4.716, p < 0.05$].

For question 4 targeting the agreement with Barbara, we did not find any main effect or any interaction regarding the distinction between fallacious and non-fallacious conditions [$F1(1,33) = 2.677, p = 0.112$; $F2(1, 39) = 1.364, p = 0.25$] and no effect either regarding the use of the connective *puisque* [$F1(1,33) = 0.6, p = 0.444$; $F2(1,39) = 3.537, p = 0.68$].

5.6 Discussion

In the second experiment, we tested the impact of the French causal connective *puisque* on the acceptability of a straw man. We hypothesized that the straw man would be better accepted when the fallacious part is juxtaposed to the previous segment rather than introduced by *puisque*, because the connective raises participants' awareness to the presence of reported content, which in our view is likely to trigger more vigilance. Our results support this hypothesis, as participants did indeed detect fallacies better when they were introduced by the connective *puisque*. More specifically, when participants read the first segment expressing a standpoint and the following connective *puisque*, they expected to read a justification for the standpoint on the one hand and to get accessible information on the other hand. In the case of the straw man, the segment following the standpoint is a misrepresentation and cannot thus be used as a justification; as a consequence, it represents new information. The discrepancy between the expectation to read given information and the conveyed new information led to a better detection of the fallacy. The chances of the straw man of remaining undetected are therefore higher when the standpoint and the fallacious argument used to justify the standpoint are not linked with a connective but juxtaposed as two separate sentences. This result thus provides important confirmation for the role of discourse connectives as procedural indicators that guide hearers in the process of utterance interpretation and demonstrate their role in the context of argumentative discourse.

These results also confirmed that fallacies are detected, as they consistently yielded lower acceptance rates compared to non-fallacious arguments, as in experiment 1. In addition, responses to questions 1 to 3 were also convergent, thus confirming further that participants used them to assess the argumentative move rather than to express their own world view, contrary to question 4.

6. Experiment 3

6.1 Participants

For this experiment, we recruited 31 French native speaking participants online (29 female, mean age: 23, age range 19-41) living in the French speaking part of Switzerland. They were all psychology students in their first year at the University of Fribourg. The participants that had already taken part in a previous experiment were excluded from participation in this experiment. Every participant received 0.25 ECTS for the participation.

6.2 Materials

We used the same design as in the previous experiments. Participants read 40 short dialogues presented in a randomized order. Again, we did not alter the first part of the dialogue (15), always expressed by Barbara, in this experiment.

- (15) Barbara : Il est crucial de mieux soutenir les jeunes parents parce qu'avoir un enfant signifie beaucoup de charges financières.
Alexandre: It is crucial to better support young parents because having a child means a lot of financial charges.

The statement expressed by Barbara remained identical in all four conditions. Like in Experiments 1 and 2, Alexandre's response to Barbara's statement contained a first segment that remained unchanged in all four conditions and a second segment that contained the manipulated variable. The first variable focused on the distinction between fallacious and non-fallacious reformulations and the second variable contained a reformulation of Barbara's argument either at the explicit or the implicit level. In the case of explicit reformulations, the wording remained almost identical to Barbara's original formulation whereas in the case of implicit reformulations, the wording differed substantially from the initial version.

The first condition presented in (16) illustrates an explicit misrepresentation of Barbara's initial argument. We created the straw man by exaggerating a lexical element of the argument. In (16), *financial charges* was modified with the more radical

noun phrase *financial ruin*. In this experiment, the exaggeration was always targeting a lexical element of the sentence but deliberately never bore on quantifier scales (e.g. some, all), in order to exclude unwanted potential effects related to the linguistic variety of items.

- (16) Alexandre : Augmentons les allocations familiales puisqu'avoir un enfant signifie la ruine financière.
Alexandre: Let's raise the family allowance since having a child means financial ruin.

By contrast, the experimental condition in (17) illustrates an implicit misrepresentation. In this condition, the wording of the second segment differs substantially from the original, and the reformulation bears on an implicit conclusion that could be drawn based on Barbara's argument.

- (17) Alexandre : Augmentons les allocations familiales puisqu'on ne pense qu'à l'argent.
Alexandre: Let's raise the family allowance since it only is about the money.

The third condition illustrated in (18) shows an explicit but non-fallacious reformulation of the original argument given by Barbara. In this case the wording remained very similar to the initial argument given by Barbara. Similarly to what we did in the fallacious condition, we rephrased one lexical part of the original argument (financial charge) in a non-fallacious way (financial weight).

- (18) Alexandre : Augmentons les allocations familiales puisqu'avoir un enfant peut être un poids financier.
Alexandre: Let's raise the family allowance since having a child can be a financial weight.

In the fourth experimental condition, illustrated in (19), we presented a non-fallacious and implicit reformulation of Barbara's argument ('having a child means a lot of financial charges'). In this condition, the wording was also different from the original ('parents are under economic pressure') but the meaning remained close to the initial argument – in both cases we put the focus on the financial aspect of parenthood.

- (19) Alexandre : Augmentons les allocations familiales puisque les parents sont sous pression économique.
Alexandre: Let's raise the family allowance since parents are under economic pressure.

As in previous experiments, experimental conditions were divided into four lists using a Latin square design, in order to ensure that every item appeared only under one condition in each list. Every participant read 10 items per condition (40 in total).

6.3 Procedure

The experiment was designed on the survey platform *Qualtrics* and distributed to the participants via weblink.

We used the same procedure as the one followed in experiments 1 and 2. The preliminary instructions, as well as the demographic questions were identical to those that were used in experiments 1 and 2. We did not place any time limit for the completion of the survey, and participants completed the task in approximately 30 minutes.

Participants responded to the same 4 questions as in experiment 1 and 2, illustrated in (6) to (9), on a 6-point Likert scale with the additional option ('I don't know').

6.4 Analysis

We performed a 2x2 repeated measure ANOVA with two within-subject factors on our data. Like in experiment 1, we targeted the nature of the argument (non-fallacious vs. fallacious). The new variable from the experiment targeted the nature of the reformulated argument: explicit or implicit.

We only included the responses given on the scale from 1 to 6 in the data analysis. We treated the responses on the additional option ("I don't know") as missing data, and removed them from the analysis.

6.5 Results

In Table 3, we report the mean score of acceptability and the standard deviation for each condition.

Table 3: explicit misrepresentation vs. implicit misrepresentation

Question 1: Proportion	Mean	Standard Deviation
Fallacious explicit	3.311	0.669
Fallacious implicit	2.868	0.742

Non-fallacious explicit	4.118	0.756
Non-fallacious implicit	3.916	0.748
Question 2: Logical Link	Mean	Standard Deviation
Fallacious explicit	3.560	0.811
Fallacious implicit	3.157	0.859
Non-fallacious explicit	4.097	0.704
Non-fallacious implicit	4.000	0.700
Question 3: Agreement with Alexander	Mean	Standard Deviation
Fallacious explicit	3.485	0.818
Fallacious implicit	3.117	0.715
Non-fallacious explicit	4.107	0.579
Non-fallacious implicit	4.145	0.671
Question 4: Agreement with Barbara	Mean	Standard Deviation
Fallacious explicit	4.458	0.623
Fallacious implicit	4.526	0.854
Non-fallacious explicit	4.622	0.623
Non-fallacious implicit	4.429	0.554

The analysis of variance returned two main effects for the first question. First, non-fallacious reformulations were better accepted than fallacious ones [$F1(1, 31) = 67.364, p < 0.001$; $F2(1, 39) = 133.640, p < 0.001$]. Second, the straw man fallacy had a higher acceptability rate when the argument was misrepresented in an explicit rather than an implicit way [$F1(1, 31) = 15.211, p < 0.001$; $F2(1, 39) = 9.705, p < 0.01$].

The first effect was reproduced for the question about the logical link between the statements. The analysis showed that participants prefer the non-fallacious conditions over the fallacious ones [$F1(1, 31) = 64.171, p < 0.001$; $F2(1, 39) = 76.621, p < 0.001$]. The analysis also indicated that the straw man was more likely to be accepted when the misrepresentation was echoing the explicit rather than the implicit content of the initial argument [$F1(1, 31) = 7.936, p < 0.01$; $F2(1, 39) = 7.268, p < 0.01$].

We report a similar tendency for the third question, targeting the agreement with Alexandre. The results confirmed that fallacious conditions were less accepted than non-fallacious ones [$F1(1, 31) = 106.494, p < 0.001$; $F2(1, 39) = 94.390, p < 0.001$]. The analysis did not return any main effect regarding the difference between explicit and implicit reformulations [$F1(1, 31) = 3.082, p = 0.890$; $F2(1, 39) = 2.636, p = 0.113$]. However, there was an interaction between the two factors, that approached significance in the participant analysis and was significant in the item analysis [$F1(1,$

31) = 3.797, $p < 0.061$; $F2 (1, 39) = 6.730$, $p < 0.013$]. An inspection of the means indicates that the agreement with Alexandre is stronger in the explicit than the implicit version of fallacious arguments, but a similar difference is not found in the case of non-fallacious arguments.

We did not find any main effect or any interaction for question 4, targeting the agreement with Barbara. Participants did not show any preference between fallacious and non-fallacious conditions ($F1 (1,33) = 0.121$, $p = 0.731$; $F2 (1,39) = 0.558$, $p = 0.459$) or implicit and explicit echoes [$F1 (1,33) = 0.865$, $p = 0.36$; $F2 (1,39) = 3.089$, $p = 0.087$].

6.6 Discussion

The third experiment aimed to assess the impact of distorting the speaker's explicit or implicit meaning. We hypothesized that the straw man fallacy would be more likely to be accepted when the distorted content was conveyed in an explicit rather than implicit way, because the explicit misrepresentation gives the illusion that the content corresponds to the original statement due to the surface similarity between the words used in the original statement and in the misrepresentation. The results tended to confirm this hypothesis. Interestingly, the interaction effect that emerged in the item analysis of the third question tends to indicate that the difference between implicit and explicit reformulations did not affect fallacious and non-fallacious arguments in a similar way. When the reformulation was not fallacious, participants judged the responses in a more similar manner than when they were based on explicit or implicit echoing of the speaker's statement. This result is in line with pragmatic theories such as relevance theory, which do not presuppose a difference in the way explicit and implicit information is understood. In both cases, the speaker's meaning is derived based on the most accessible information, be it implicit or explicit, by a combination of decoding and inference. Given that this effect was significant only in the analysis by items and only approaching significance in the analysis by participants, it will need to be replicated in future research before strong conclusions can be reached in this respect. Our data indicates, however, that this could be a promising line of inquiry.

We assume that participants' blindness to the distortion of the speaker's meaning found in the explicit straw man is akin to a form of semantic illusion. Because discourse processing is shallow, a surface form similarity is sometimes mistaken for a fully faithful

echo of the speaker's utterance. In the implicit version, by contrast, the link between speaker meaning and the paraphrase requires an inferential process on the part of the hearer. Such process is costly and might involve a deeper form of processing, thus leading to a better detection of meaning differences between the original and the reported statements, which in turn would lead to higher detection rates.

7. General Discussion

In this paper, we presented three experimental studies testing the acceptability of the straw man fallacy. It appears from our results that the acceptability of the straw man fallacy can indeed be influenced by linguistic factors, as all three experimental manipulations yielded significant differences in the acceptability rates for straw men fallacies. More specifically, our results demonstrate that the locus of the fallacious argument has an impact on its acceptability. Results from Experiment 1 indicated that misrepresentations of arguments represent more acceptable cases of straw men compared to misrepresented standpoints. In addition, the use of the connective *puisque* also has an effect on the acceptability of the straw man. Results from Experiment 2 indicated that introducing a fallacious argument by using a causal connective with an echoic meaning raises participants' awareness to the possibility of a fallacious argument. Finally, the level of pragmatic enrichment involved in comparing the speaker's initial statement and its attributed version also affect participants' detection of straw man fallacies. Results from Experiment 3 indicated that a reformulation involving the speaker's explicit content leads to more acceptable straw men, due to the illusion of similarity that they create. In sum, according to our data, effective straw men target the opponent's argument, do not make use of explicit connectives and echo their opponent's explicit meaning.

These results are important for argumentative scholarship on the straw man fallacy on a number of counts: first of all, and this is significant from a theoretical perspective, they suggest that the pragma-dialectical characterization of the straw man fallacy as a fallacy violating the standpoint rule needs to be adapted, since our results show that participants are less fooled by straw men on standpoints than straw men on arguments. This additionally testifies to the actual difference in status of the two argumentative moves, which might also have further theoretical implications, for instance in terms of potential specifications and qualifications to accounts such as

Pollock's (1987). Second, while we by no means claim to have conducted an exhaustive investigation of all possible linguistic resources used for misrepresentation, we now have supplied experimental evidence that (i) the use of explicit connectives jeopardizes the success of the fallacy and that (ii) close renditions (restricted to lexical items) of the original utterance are more likely to succeed. Moreover, we provided cognitively-grounded assumptions as to why this is the case. In this respect, our results have to be construed as a genuine descriptive and explanatory contribution to the research on argumentative indicators and, more generally, to the blooming research on the links between argumentative practices and their linguistic realization (see e.g., Oswald et al. 2018). Finally, the third experiment highlights the relevance of a (cognitive) pragmatic perspective on the study of argumentative practices which goes beyond the classical speech-act-theoretic considerations typically found in the field.

Even though our experiments represent an important step ahead to empirically assess the factors influencing the acceptability of straw man fallacies, it is clear that other factors that we have not tested yet might also influence their acceptability. First, our experimental design involved written dialogues. Yet, in oral argumentative contexts, other features such as the speaker's tone of voice, facial expressions, accent, etc. also contribute to their perceived reliability and are likely to influence the acceptability of fallacies. The role of these features in an oral setting will have to be analyzed in future work. The dialogues of our experiments perfectly lend themselves to an oral experimental design and could be reused to further assess their validity.

Because of the constraints of experimental testing, we only included cases of straw men that followed a strict form of linguistic formulation, with a standpoint immediately followed by only one argument. It is clear that in real discourse situations, straw man fallacies can come into an array or variations on this format. Based on our results, we cannot claim that the effects we uncovered with the factors that we tested would be replicated in all other formats. The difference between the efficiency of distorted standpoints vs. arguments in particular might vary in situations in which the standpoint is more linguistically distant from the fallacious reformulation, or in cases in which several arguments are mentioned to back-up a standpoint. But then again, these factors also lend themselves to empirical testing and could be the focus of future experiments. Similarly, the difference between explicit and implicit forms of reformulations could be affected by their distance with the source that is echoed.

Indeed, this factor has been shown to affect people's ability to detect cases of semantic illusions (Sanford, 2002). However, longer distances often imply a decreased rather than an increased ability to spot the semantic mistake, we can therefore expect that creating a longer distance would increase rather than decrease the acceptability of the straw man.

Another important finding from our experiments concerns the role of the connective *puisque* as a marker of attributed content. This connective seems, however, to be highly language specific, as *puisque* was found not to have exact translation equivalents in other languages like Dutch (Degand, 2004), German (Pit, 2007) or English (Zufferey & Cartoni, 2012). It is therefore an open question if and whether other causal connectives in other languages would play a similar role. In English, connectives like *since* (van Eemeren et al., 2007) and *after all* have also been described as indicators of given information (Ariel, 1985) and could also be tested in argumentative contexts.

In our experiments, we assessed participants' detection of straw man fallacies using three different formulations (plus a control question). In most cases, all formulations yielded a similar pattern of results, but some differences also emerged, especially in Experiment 3. In our experimental settings, the three questions were always presented in the same order. It is noticeable that the most reliable effects were found on our first question, targeting the exaggerated nature of the straw man fallacy. In future experiments, it would be interesting to vary the order of presentation of the questions in order to assess whether it is really the exaggerative nature of the straw man that participants detect better than the logical aspect, or whether this effect was due to the fact that this was the first question that they had to answer. Similarly, it would be interesting to test other question formulations to assess participants' evaluation of the straw man fallacy. Then again, these formulations only assess participants' conscious evaluation of the fallacy rather than their intuitive reactions to it. Further experiments using online reading should be used to complement these results with data about online processing. Indeed, offline judgment and online processing have often been found to yield different patterns of results in other domains of linguistic competence, and it would be interesting to assess both aspects in the case of argumentative dialogues.

In sum, the three experiments reported in this paper represent a first attempt to empirically assess the role of linguistic factors for the acceptability of the straw man fallacy. It opens a new avenue of inquiry that can and will have to be complemented by further studies to assess the role of other factors in both written and spoken modality, across a range of different types of fallacies.

8. References

- Aikin, Scott F. & Casey, John F., 2011. Straw man, weak man and hollow man. *Argumentation* 25, 87-105. <https://doi.org/10.1007/s10503-010-9199-y>
- Aikin, Scott F. & Casey, John F., 2016. Straw man, iron man and argumentative virtue. *Argumentation* 35: 431-440. <https://doi.org/10.1007/s11245-015-9308-5>
- Ariel, Mira, 1985. The discourse functions of given information *Theoretical Linguistics* 12, 99-113.
- Bizer, George Y., Kozak, Shirel M. & Holterman, Leigh Ann, 2009. The persuasiveness of the straw man rhetorical technique. *Social Influence* 4 (3), 216-230. <https://doi.org/10.1080/15534510802598152>
- Blakemore, Diane, 1987. *Semantic constraints on relevance*. Oxford, UK: Blackwell.
- Blakemore, Diane, 2002. *Relevance and Linguistic Meaning. The semantics and pragmatics of discourse markers*. Cambridge, UK: Cambridge University Press.
- Bohan, Jason, 2008. *Depth of processing and semantic anomalies (Doctoral thesis)*. Retrieved from <http://theses.gla.ac.uk/127/1/2008bohanphd.pdf>
- Caron, Jean, Micko Hans Christoph & Thüring, Manfred, 1988. Conjunctions and the recall of composite sentences. *Journal of Memory and Language*, 27 (3), 309-323. [https://doi.org/10.1016/0749-596X\(88\)90057-5](https://doi.org/10.1016/0749-596X(88)90057-5)
- Degand, Liesbeth, 2004. Contrastive analyses, translation, and speaker involvement: the case of *puisque* and *aangezien*. In: Achard, Michel & Kemmer, Suzanne (eds.), *Language, culture and mind*. Stanford, Canada: CSLI Publications, pp. 251-270.
- van Eemeren, Frans H., Garssen, Bart, Krabbe, Erik C. W., Snoeck Henkemans, A. Francisca, Verheij, Bart & Wagemans, Jean H. M., 2014: *Handbook of argumentation theory*. Dordrecht, Netherlands: Springer.
- van Eemeren, Frans, H., Garssen Bart & Meuffels, Bert, 2012. The disguised abusive ad hominem empirically investigated. *Strategic manoeuvring with direct personal attacks*. *Thinking & Reasoning* 18 (3), 344-364. <https://doi.org/10.1080/13546783.2012.678666>
- van Eemeren, Frans H. & Grootendorst, Rob, 1992. *Argumentation, communication, and fallacies: A pragma-dialectical perspective*. Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.
- van Eemeren, Frans H. & Grootendorst, Rob, 2004. *A systematic theory of argumentation*. Cambridge, UK: Cambridge University Press.
- van Eemeren, Frans H., Grootendorst, Rob & Snoeck Henkemans, A. Francisca, 2002: *Argumentation. Analysis, evaluation, presentation*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- van Eemeren, Frans H. & Houtlosser, Peter, 1999: *Strategic manoeuvring in argumentative discourse*. *Discourse Studies* 1 (4), 479-497.
- van Eemeren, Frans H., Houtlosser, Peter & Snoeck Henkemans, A. Francisca, 2007. *Argumentative indicators in discourse. A pragma-dialectical study*. Dordrecht, Netherlands: Springer.
- Erickson, Thomas D. & Mattson, Mark E., 1981. From words to meaning: a semantic illusion. *Journal of Verbal Learning and Behaviour* 20 (5), 540-551. [https://doi.org/10.1016/S0022-5371\(81\)90165-1](https://doi.org/10.1016/S0022-5371(81)90165-1)
- Franken, Nathalie, 1996. Pour une nouvelle description de *puisque*. *Revue Romane* 31 (1), 1-18.
- Hamblin, Charles L., 1970. *Fallacies*. London, UK: Methuen.
- Hansen, Hans V. & Pinto, Robert C. (eds.), 1995. *Fallacies: classical and contemporary readings*. University Park, PA: The Pennsylvania State University Press.
- Lewiński, Marcin, 2011. Towards a critique friendly approach to the straw man fallacy evaluation. *Argumentation* 25, 469-497. <https://doi.org/10.1007/s10503-011-9227-6>

- Lewiński, Marcin & Oswald, Steve, 2013. When and how do we deal with straw men? A normative and cognitive pragmatic account. In Maillat, Didier & Oswald, Steve (Eds.). *Biases and constraints in communication: argumentation, persuasion and manipulation*. Special issue of the *Journal of Pragmatics*. <https://doi.org/10.1016/j.pragma.2013.05.001>
- Lillo-Unglaube, Marco, Canales-Johnson, Andrés, Navarrete, Gorka & Fuentes Bravo, Claudio, 2014. Toward an experiment account of argumentation: the case of the slippery slope and the ad hominem arguments. *Frontiers in Psychology*, 5. <https://doi.org/10.3389/fpsyg.2014.01420>
- Macagno, Fabrizio & Walton, Douglas, 2017. *Interpreting straw man argumentation. The pragmatics of quotation and reporting*. Cham: Springer International Publishing. <https://doi.org/10.1007/978-3-319-62545-4>
- Millis, Keith K., Golding, Jonathan M. & Barker, Gregory, 1995. Causal connectives increase inference generation. *Discourse Processes* 20, 29-49. <https://doi.org/10.1080/01638539509544930>
- Morency, Patrick, Oswald, Steve & Saussure, Louis de, 2008. Explicitness, implicitness and commitment attribution: a cognitive pragmatic approach. *Belgian Journal of Linguistics* 22, 197-219. <https://doi.org/10.1075/bjl.22.10mor>
- Oswald, Steve & Lewiński, Marcin, 2014. Pragmatics, cognitive heuristics and the straw man fallacy. In Herman, Thierry & Oswald, Steve (Eds.). *Rhétorique et cognition: perspectives théoriques et stratégies persuasives*. Bern, Switzerland: Peter Lang, pp. 313-343.
- Oswald, Steve, Herman, Thierry & Jacquin, Jérôme, 2018. *Argumentation and Language: Linguistic, Cognitive and Discursive Explorations*. Cham: Springer.
- Ozols, Davis, Maillat, Didier & Oswald, Steve, 2016. Repetition as a context selection constraint: A study in the cognitive underpinnings of persuasion. In: Mohammed, D. & Lewiński, M., *Argumentation and Reasoned Action, Proceedings of the 1st European Conference on Argumentation, Lisbon 2015*. Volume 1, pp. 547-566.
- Pit, Mirna, 2007. Cross-linguistic analyses of backward causal connectives in Dutch, German and French. *Languages in Contrast* 7, 53-82.
- Pollock, John L., 1987. Defeasible Reasoning. *Cognitive Science* 11, 481-518.
- Sanders, Ted J. M. & Noordman, Leo G. M., 2000. The role of coherence relations and their linguistic markers in text processing. *Discourse Processes* 29, 37-60. https://doi.org/10.1207/S15326950dp2901_3
- Sanford, Alison J. S., Price, Jessica & Sanford, Anthony J., 2009. Enhancement and suppression effects resulting from information structuring in sentences. *Memory & Cognition* 37 (6), 880-888. <https://doi.org/10.3758/MC.37.6.880>
- Sanford, Alison J. S., Sanford, Anthony J., Molle, Jo & Emmott, Catherine, 2006. Shallow processing and attention capture in written and spoken discourse. *Discourse Processes* 42 (2), 109-130. https://doi.org/10.1207/s15326950dp4202_2
- Sanford, Anthony J., 2002. Context, attention and depth of processing during interpretation. *Mind & Language* 17 (1-2), 188-206. <https://doi.org/10.1111/1468-0017.00195>
- Sanford, Anthony J. & Graesser, Arthur C., 2006. Shallow processing and underspecification. *Discourse Processes* 42(2), 99-108. https://doi.org/10.1207/s15326950dp4202_1
- Sanford, Anthony J. & Sturt, Patrick, 2002. Depth of processing in language comprehension: not noticing the evidence. *TRENDS in Cognitive Sciences* 6 (9), 382-386. [https://doi.org/10.1016/S1364-6613\(02\)01958-7](https://doi.org/10.1016/S1364-6613(02)01958-7)
- Saussure, Louis de, 2018. The straw man fallacy as a prestige-gaining device. In Oswald, Steve, Herman Thierry & Jacquin, Jérôme (Eds.). *Argumentation and Language – Linguistic, Cognitive and Discursive Explorations*. Dordrecht, Netherlands: Springer, pp. 171-190.
- Sperber, Dan & Wilson, Deirdre, 1995. *Relevance. Communication & Cognition*, 2nd edition. London, UK: Blackwell.
- Sperber, Dan, Clément, Fabrice, Heintz, Christophe, Mascaro, Olivier, Mercier, Hugo, Origgi Gloria & Wilson, Deirdre, 2010. Epistemic Vigilance. *Mind & Language* 25 (4), 359-393.
- Tindale, Christopher, 2007. *Fallacies and argument appraisal*. Cambridge, UK: Cambridge University Press.
- Uzelgun, Mehmet A., Mohammed, Dima, Lewiński, Marcin & Castro, Paula, 2015. Managing disagreement through yes, but... constructions: an argumentative analysis. *Discourse Studies*, 17 (4), 467-484. <https://doi.org/10.1177/1461445615578965>
- Walton, Douglas, 1995. *A pragmatic theory of fallacy*. Tuscaloosa, AL: The University of Alabama Press.
- Wang, Lin, Hagoort, Peter & Yang, Yufang, 2009. Semantic illusion depends on information structure: ERP evidence. *Brain Research* 1282, 50-56. <https://doi.org/10.1016/j.brainres.2009.05.069>

- Zufferey, Sandrine, 2012. *Car, parce que et puisque* revisited: three empirical studies on French causal connectives. *Journal of Pragmatics* 44, 138-153. <https://doi.org/10.1016/j.pragma.2011.09.018>
- Zufferey, Sandrine, 2014. Givenness, procedural meaning and connectives. The case of French *puisque*. *Journal of Pragmatics* 62, 121-135. <https://doi.org/10.1016/j.pragma.2013.09.022>
- Zufferey, Sandrine & Cartoni, Bruno, 2012. English and French causal connectives in contrast. *Languages in Contrast* 12 (2), 232-250. <https://doi.org/10.1075/lic.12.2.06zuf>