Chapter 8
The Straw Man Fallacy as a Prestige-Gaining Device

Louis de Saussure

Abstract In this paper, we consider the straw man fallacy from the perspective of pragmatic inference. Our main claim is that the straw man fallacy is a 'pragmatic winner' not primarily because of its persuasive power but rather because it targets the pragmatic cognitive-inferential skills of its victim while enhancing the prestige of its author. We consider that in the context of a straw man fallacy, the issue of the burden of proof, which is 'reversed', does not directly bear on the argumentation itself but has essentially to do with the difficulty for the targeted speaker of getting the attention of the audience back. It is difficult because countering this fallacy involves primarily a discussion of the reasons why the inference drawn (the meaning or the thought fallaciously attributed to the targeted speaker) was unduly derived, a process which is virtually destined to be a failure first of all because of the lack of relevance (in the sense of Sperber and Wilson in Relevance. Communication and cognition. Blackwell, Oxford, 1995) of *justifications* in comparison with that of actual points. Notions of retractability and the explicit-implicit divide are central to our approach.

8.1 Introduction[1]

Among the main questions raised by fallacies, these two are of particular interest not only for the argumentation scholar but also to the pragmatist, if considering the notion of 'winning an argument' from a wide, pragmatic and communicative, perspective:

[1]This research is part of a project funded by the Swiss National Science Foundation (project N° 100015_146751, *Biased Communication: the Cognitive Pragmatics of Fallacies*).

L. de Saussure (✉)
Institut des Sciences du Langage et de la Communication, Université de Neuchâtel, Neuchâtel, Switzerland
e-mail: louis.desaussure@unine.ch

S. Oswald et al. (eds.), *Argumentation and Language-Linguistic, Cognitive and Discursive Explorations*, Argumentation Library 32,
https://doi.org/10.1007/978-3-319-73972-4_8

(1) How is it that fallacies are successful in persuading?
(2) How is it that fallacies may occasionally fail in persuading an audience but still constitute a winning move in some argumentative interactions?

The description of fallacies and of their effects does not exhaust the need for deeper, cognitive, explanations relatively to their actual efficiency on these two levels.

The first question is about *how* we actually process information: since fallacies are demonstrably invalid or unacceptable from a normative point of view, they have to be processed through 'peripheral' routes (cf. Petty and Cacioppo 1986), relying on heuristics (Tversky and Kahneman 1974) and 'fast and frugal' processes (Gigerenzer 2004).[2]

Human verbal information processing is not only about comprehending utterances or grasping others' acts of communication. Information processing does not only have an informational outcome, but also an epistemic one. These processes have a bearing on epistemic attitudes and as such they involve some sort of evaluation. Fallacies, just as other types of arguments, are filtered out by our cognitive abilities of 'epistemic vigilance' (Sperber et al. 2010) before they are integrated in the cognitive environment (i.e. before they are accepted as true).

The second question has to do with what humans actually do with information as far as actual individual opponents are considered. The literature has various views on the question of why we actually enter into argumentative dialogues. The first hypothesis that comes to mind is that we do so in order to achieve better knowledge through an informed exchange of ideas. It's certainly generally true but arguing involves several layers of problems and several reasons to be persuaded or convinced. Actually, this question has two aspects, reflected in the two questions above. First, one can be impressed by an argument. Second, however, one can be impressed by the argumentative skills of an individual, regardless of the quality of the arguments themselves.

Dessalles (2011) develops a theory where language has basically two functions, narration and argumentation. Argumentation is then viewed as being only incidentally an ability to do collective problem-solving; its main function is rather to chase the lies of others, which entails gaining prestige out of it and promoting oneself in the social hierarchies and patterns of dominance against others. A related claim is made by Pinker et al. (2008) in a paper on how implicit meaning, therefore retractability, enables the passing of proposals to others without being formally held responsible for them (we will discuss more precisely the question of retractability further down). They claim that the fact that speakers may retract from meanings to which they are not formally committed enables them to enter complex interactions

[2]These approaches are not similar; they diverge on various points which we cannot discuss here. They however all tackle the fact believability often relies on other processes than analytical, reflexive, judgement.

where briberies, and other forms of manipulative attempts, can be performed without being exposed to social sanctions.[3]

Commitment (in the sense of Hamblin 1970: the contents to which the speaker is formally committed to having put forward),[4] and therefore retractability, are logically binary: a speaker cannot retract from what is said (unless in order to make a corrective statement), because retracting from what is overtly stated raises a formal inconsistency. On the contrary, what is implicated is retractable.[5] However, de Saussure and Oswald (2009) point out that in fact, there are indeed implicit meanings that are, pragmatically, very unlikely to be retracted without a feeling of bad faith (that is a sense of informal, or pragmatic, inconsistency). Logical inconsistency is of course a true-or-false objective property, but retractability is actually a gradual psychological notion.

Contents which are communicated pragmatically can be more or less committing in contexts, but all are logically cancellable without formal inconsistency. As a principle, implicatures that are remote from the linguistic form are perhaps more likely to be retracted without problems, and pragmatic modulations that are mere extensions of the linguistic form, or 'explicatures' (cf. Sperber and Wilson 1995), are more likely to trigger feelings of bad faith if they are retracted, but these are only main trends. De Saussure and Oswald (2009) claim that the more a content is relevant in the circumstances, the less it can be retracted without raising issues of bad faith. For this reason, inferences that come up automatically in most contexts, although they are of course available or cancellable in specific contexts, are unlikely to allow for a retraction without raising issues of bad faith. An inference like 'Mary has exactly four children' derived from utterance *Mary has four children*, or a scalar implicature like 'Not all students came to the party' from *Some students came to the party*, or even a conventionalized indirect speech act like 'I ask you to pass me the salt' triggered by utterance *Can you pass the salt?* are strongly committing, i.e. they can't be retracted without raising a feeling of bad faith. Needless to say, such processes concern potentially all types of utterances and all types of contexts, argumentation and persuasion among them, fallacious argumentation included.

Making a strawman fallacy (henceforth SMF) is precisely a way of playing with these elements, since it involves processes such as extracting pragmatic meanings, attributing thoughts and intentions on the basis of behaviour, and claiming that what is in principle retractable is on the contrary obviously unretractable.

[3]For a thorough discussion of language evolution at large, including these types of verbal behaviour, see Reboul (forthcoming 2017).

[4]See Morency et al. (2008) and de Saussure and Oswald (2009) for elaborations on the notion of commitment and how it can be dealt with from a pragmatic perspective.

[5]Needless to say, the notion of retractability is the cognitive psychological counterpart of the formal cancellability of implicatures pointed out by H. P. Grice. We will come back to these notions below.

In this chapter, we suggest that the processes involved in the search for relevance (following the notion by Sperber and Wilson 1995)[6] are a key factor in the success of fallacies at two levels: (i) they prompt the feeling, by individuals, of their acceptability despite objective logical deficiency, and (ii) the authors of fallacies manifest their own ability in deriving inferences and predicting relevant pieces of meaning, through their rhetorical ability, which enables them to gain prestige and the allegiance of an audience while disrupting their epistemic vigilance.

We suggest that the SMF is particularly efficient in manifesting a superior rhetorical ability; the SMF is a winning argument, and it is so for pragmatic reasons, since it exploits and satisfies particularly well the expectations of economy raised by the hearers—in particular because SMFs arise virtually only in order to gain the support of a third party (a general audience, in general) and thus to gain a dominating position in the social context.

In the following section, we will start by exposing a pragmatic puzzle with the straw man fallacy: whereas it could simply be an obvious misunderstanding, given the pragmatic principles of interpretation, the SMF appears on the contrary as a particularly successful interpretation. Then, in Sect. 8.3, we argue that the straw man fallacy primarily targets not the meanings but the pragmatic abilities of the speaker. Section 8.4 addresses more specifically the issue of the *burden of proof*.

8.2 A Pragmatic Puzzle with the Straw Man Fallacy

8.2.1 The Straw Man Fallacy: Properties and Problems

The SMF consists in attributing to an individual, generally on the basis of her verbal utterances,[7] commitments to contents which she did not actually intend to convey. The literature offers many examples of SMFs constructed for the aim of the theoretical understanding of the notion, such as these, taken from Lewiński and Oswald (2013):

(3) A: Many right-wing politicians are devout believers.
B: I am not so sure that most right-wing politicians are devout believers.

[6]Sperber and Wilson (1995, first edition 1986) argue that the process of comprehending an utterance relies on a principle which can be summarized as follows: 'look for relevance'. The hearer assumes (and the speaker assumes that the hearer assumes) that the speaker is speaking relevantly: his utterance provides as much information as expected for the smallest possible cognitive expense.

[7]In fact, the SMF may also occur on the basis of other communicative or even non communicative behaviours, which is not surprising because this fallacy is basically about attributing intentions and thoughts to an individual, which can be reconstructed on the basis of various forms of behaviour, not only verbal utterances. This chapter focuses specifically on those SMFs that are linked to verbal utterances.

(4) A: Social policies of the government are plainly inefficient: a number of scientific studies, including one recently published in *Sociology*, expose major faults of the policies.
B: It's funny to say that the government's social policies are inefficient based on just one scientific study.

(5) A: In fact, the majority voted in favour, but the motion was not accepted since there was no quorum needed for the occasion.
B: I'm sad to hear the majority rule does not apply to our parliament anymore!

In short, what the arguments B here above do is simply to attack the speaker A on a position that she did not actually hold. Here is an authentic example of a straw man fallacy: during a TV debate, one of the interlocutors (French comedian Ramzy Bédia), annoyed by his opponent (French polemist Eric Zemmour) who constantly makes literary quotations, complains[8]:

(6) Ramzy Bédia: Can't we speak normally without dropping names on every occasion?
Eric Zemmour: Excuse me for having read books.

(In the case above, we observe something like a shift from complaining about too profusely showing off one's knowledge to complaining about simply having knowledge at all.)

In principle, we know what fallacies are: they are arguments which depart from norms of valid or acceptable reasoning in some formal or pragmatic way. What is surprising is that humans have at the same time two opposite properties in this respect: they are able to distinguish fallacious from sound arguments upon reflection, and are prone to be persuaded by fallacious arguments. Yet, it is also true that a number of argumentative schemes can be fallacious in some contexts but sound in others, such as the argument of authority or the *ad populum*. We also know that cognitive processes can take various heuristic routes, therefore they bypass critical evaluation and lead to systematic results, which allow suggesting that the success of fallacious arguments reside in the exploitation of such heuristics; yet, in turn, these heuristics prove very useful in ordinary, non-malevolent, contexts.

All in all, there is a true complexity in the project of identifying fallacious arguments and explaining their success. In particular, even though there is abundant literature about cognitive biases on one side and about fallacies on the other, we have little knowledge of how specific types of fallacies—for example the SMF—work precisely. There are also a number of cognitive approaches to argumentation and manipulation which provide fundamental lines along which explanations can be brought, to which we will turn further down.

[8]Example by courtesy of Thierry Herman, Steve Oswald and Misha Müller. Our (approximate) translation.

The SMF involves an original speaker (henceforth 'the speaker') who utters some linguistic string, and an opponent who makes a SMF [henceforth 'the author' (of the SMF)]. It has four key properties which, once put together, are puzzling.

First, the SMF is about reversing the burden of proof to the (original) speaker. Interestingly, while in many situations the author of a questionable interpretation is expected to make justifications about it, in the case of the SMF, the opposite is happening: the author of a SMF is making an incorrect interpretation of the speaker's utterance but it is the speaker, not the author of the SMF, who finds herself with the burden of proving that the interpretation is incorrect. A problem with the reversal of the burden of proof is that it impairs the speaker's performance in the argumentative interaction (it is a 'burden'). After all, there is no obvious reason for which *providing justifications* should be an impairment, a 'burden', in an argumentation: it could as well be an opening to furthering the discussion and openly rectifying the misinterpretations. But it is not.

The second remarkable property of SMFs is that they generally involve a (silent) third-party witnessing the exchange and whose support is important to the interlocutors.

Hence a question: does the third party have to be persuaded by the SMF or not for the SMF to be a winning move in a debate? The question is not absurd: after all, an incorrect interpretation may actually have little persuasive power. In fact, we even venture to suggest that in some cases, if not in most cases, the external audience is indeed not persuaded by the fairness and soundness of the inference leading to the SMF. And if we are correct, then how come the speaker, but not the author of the fallacy, falls short of argumentative resources and usually surrenders in front of the reversal of the burden of proof? And how come this reversal happens at all, if the audience is not persuaded itself? In other words, how is it that attributing false conclusions to a speaker does not ridicule the author of the wrong conclusions but places the burden of proof on the shoulders of the speaker who did not intend them? And this, moreover, even when there is no acknowledgement of the validity of the interpretation on the part of the audience?

The third property, in line with what precedes, is that the SMF fails (of course) to persuade the targeted speaker, since no one can be persuaded that she meant something or that she thinks something.

The fourth property is the key to a puzzle and we will address it in the following section.

8.2.2 The Puzzle with Straw Man Fallacies

We are now approaching a funny conclusion: the SMF is an argument which is not raised in order to persuade the addressee, and which may well even fail in persuading the witnessing audience, and still be a winning move in an argumentative dialogue.

We noticed above that the SMF generally involves a third party. However, there are also cases of SMFs in face-to-face interactions without involving a third party. This is important to underline: because of this, and because no one can be persuaded that he entertains some thought or has some meaning intentions, therefore, it can simply *not* be about persuading, or at least, it has other, more fundamental, effects, than persuading.

The fourth property of the SMF is that it relies on pragmatic processing (*cf.* de Saussure and Oswald 2009; Lewiński and Oswald 2013; Oswald and Lewiński 2014). The SMF relies on pragmatic processing simply because the alleged standpoint has never been explicitly stated and therefore it is a pragmatic reconstruction on the basis of what is said (or, better, on the basis of what is explicitly communicated).[9] It targets an unsaid element, i.e. something that counts as a pragmatic inference. Therefore, the SMF is obtained by following a path of inference and thus is derived from the original utterance by following pragmatic principles.

Altogether, these properties lead to a puzzle which unfolds as follows:

(a) The SMF is pragmatic because it targets an unsaid element and therefore it exploits standard pragmatic meaning derivation procedures.
(b) Communication is generally efficient, in particular because, usually, speakers' predictions and hearers' inferences match one another in communication, which, generally and in fact virtually always, involves pragmatic meaning discovery procedures. Of course, misunderstandings do occasionally occur, but usually the hearers get the right pragmatic meaning intended by the speakers because a speaker makes successful predictions about the ability of the hearer to contextualize appropriately the utterance so that the intended meaning, including in particular what was not properly *said*, is recovered safely enough.
(c) Given (a) and (b), it is expectable that the author of a SMF manifestly displays a misunderstanding, therefore a failure. But instead of that, the SMF is not taken as a misunderstanding but as a particularly witty understanding, which goes far beyond the speaker's ostensive meaning intention, and which even reverses the burden of proof.

This is a pragmatic puzzle, in the sense that it is not a question of semantic decoding but a question of pragmatic language use, inferences and contexts.

In what follows, we suggest that this puzzle cannot be solved if we fail to take into account elements which are in some way beyond meaning itself.

[9]Even if this reconstruction may come from some manipulation of 'what is said': such a manipulation is actually extracting a potential inference (for example, when *many* is rephrased as *all* as an inference of an understatement).

8.3 The Straw Man Fallacy Targets Skills, Not Meanings

8.3.1 Cancellations and Retractions

Pragmatic meaning is pervasive in verbal communication; it develops into various levels and types of contents, from indirect speech acts to non-literal communication and to presupposing a shared common ground. These dimensions of meaning all have in common the fact that they are inferential in some way: these meanings are not directly decoded from the sentence but are constructed on top of them in order to form a plausible assumption about what the speaker seeks to render manifest to the audience.

Whereas Grice's theory of implicature viewed the world of pragmatic meaning as one big category of 'what is meant by the speaker' (as opposed to 'what is said by the sentence'), the main trends in semantics and pragmatics today consider that such inferential meanings can be classified (at least) in the two separate dimensions that we mentioned in the introduction: implicatures 'proper' and explicatures, which are these elements of meaning which are context-dependent extensions of the linguistic 'logical' form, i.e. add-ons or precisions to what was actually verbalized.

A very relevant aspect of the distinction between implicatures and explicatures in our discussion of the SMF relates to speaker's commitments on contents and their cancellability, as mentioned in the introduction. Let us shortly elaborate on this issue.

Pragmatic contents are, by nature, cancellable without creating a formal inconsistency. This applies to conversational implicature as a principle and is even a test to identify them in the classical Gricean approach. When uttering *It's raining*, the Speaker can easily control the interpretation of unwanted implicatures (such as *you should not go out*) by preventing the hearer from deriving them, by uttering something like *but you can play tennis anyway*. Similarly, the speaker of *Mary has four children*, which is normally triggering the explicature *Mary has exactly four children,*[10] can then utter *she even has five*, and there is no formal contradiction.

There are other pragmatic contents which similarly are cancellable: weak implicatures, i.e. these inferences which are not necessarily part of the intentional meaning but which are rational consequences of them, such as *the speaker doesn't like luxury* from a sentence like "I don't like fancy cars", can be controlled just as any other implicature. Equally, elements of the common ground may occur to a hearer in the course of interpreting (presupposed types of weak implicatures, or *discursive presuppositions* in de Saussure 2013); they are also cancellable.

The cognitive corollary of cancellability, in conversation, is *retractability*.

If implicit meanings are cancellable, then it follows that the speaker is not formally committed to them: they are constructed, deduced, derived, by the hearer

[10]This is a scalar inference. Such inferences are sometimes still named *scalar implicatures* by reference to the Gricean conception that there is only 'what is said' and 'what is implicated' but they actually fall within what is felt as explicit but context-dependent.

and are under his responsibility: it is the hearer who makes the (generally spontaneous) decision to use this or that piece of knowledge, present in his cognitive environment, as an implicit premise.[11]

If a wrong inference is made, i.e., if the speaker did not correctly predict how comprehension processes would be worked out by the hearer, she can make manifest that something was not intended by uttering a correction, for example by saying *that's not what I meant* in order to retract from the pragmatic content incorrectly constructed by the hearer. This is not a cancellation, since a cancellation occurs immediately, because the speaker correctly predicts the possibility of the inference. It's a retraction, because the speaker makes it *ex-post*, only when she realizes that the unintended inference was actually drawn.

But of course, only the speaker knows what she meant, and elaborating a formal proof of why this or that was not pragmatically meant is a challenge which can probably never be met. It is part of the game that if a speaker clarifies that she did not mean something, we accept it as a fact: asking for virtually impossible proofs would seem clearly unfair.

Such a corrective utterance would also be the natural reaction of a speaker in front of a SMF when she discovers the misinterpretation that has occurred to her interlocutor.

Retractability is not similar to cancellability. Whereas cancellability is a binary property of propositions relatively to linguistic forms (content is or is not cancellable), retractability is a psychological attitude of humans about their responsibility in communicating and interpreting what was not said. Cancellability can be objectively observed by a logical analysis but retractability is primarily a matter of intuitive feeling.

Retractability is therefore not subject to a binary evaluation: contents are more or less retractable depending on the feeling of people in front of utterances, interpretations, and on their own confidence in what the speaker should expect from her speech acts. Retractions are of the style 'I did not mean it' and as such are not valid or invalid: rather, they are felt as fair or unfair to various possible degrees. In sum, retractability is more or less *plausible*, whereas cancellation is or is not *valid*.

Cancellability and retractability can of course very much converge: an implicature such as *you can't go and play tennis now* from an utterance like *It's raining* is cancellable, of course, and in most contexts the speaker can undisputedly retract from it.

However the two can very much diverge too. In de Saussure and Oswald (2009) we argued that in some cases, a formal cancellation can be validly performed, i.e. without any logical inconsistency, but at the same time, a retraction will leave the hearer with a feeling of bad faith. This happens with all pragmatic contents: implicatures, strong or weak, explicatures, and background assumptions.

[11]For more details on this process, as well as on the notions of strong and weak implicatures and explicatures, see Sperber and Wilson (1995).

With an implicature, since we are dealing with conclusions drawn through non-demonstrative inferences, obviously, the feeling of bad faith in front of a retraction is usually less blatant than with explicatures. But suppose a father is telling his son: "There are a lot of people at your wedding party" (de Saussure and Oswald 2009). Given the triviality of the explicit piece of information (it is indeed expected that there is a lot of people at a wedding party) the hearer might experience a sense of lack of informativeness, leading, for the sake of recovering a relevant piece of information, to further inferences about what the speaker has in mind. Add to this some contextual information such as the fact that the father is paying the bill, and you get the derivation of a complaint about the costs, that is, an implicature like *it's going to be very expensive*. The son should be expected by the father to get this implicature and he may either swallow it or on the contrary react by something like "Dad, stop always complaining about money". To this, in turn, the father can respond that he did not imply anything. But the feeling of bad faith is expectably going to be strong, given the triviality of what was said and the context of paying the bill, even though the cancellation behind it is valid and possibly blocks further discussions.

This example directs us towards the following assumption: the most relevant inference in the circumstances is the more difficult to retract without bad faith. By 'more *relevant*' we mean, following Sperber and Wilson (1995), an inference that provides the most optimal balance between processing costs (the further remote the inference is, that is, the more derivations and incorporations of implicit premises it involves in the deduction, the costlier it is to process for the mind) and rewarding cognitive effects (a piece of information that has more consequences on the assumptions held by the hearer has more cognitive effects). In the case above, it is the lack of such cognitive effects in the literal, explicit, meaning, which triggers further inferences, imposing a supplementary cost but an informational reward when the hearer speculates that the speaker means something about *too many people/too expensive*.

More precisely, we suggest that among the possible inferences that can be derived from the utterance in given circumstances, the more relevant one is expectedly the most likely to be recovered by the hearer. If this is a general pragmatic principle of communication using mindreading (theory of mind) abilities, then the hearer naturally expects the speaker to be aware of the likeliness of its derivation. If the speaker does not prevent it by anticipating a cancellation, then it is legitimate to assume that it is going to be derived and the speaker is intuitively—not formally—considered aware of conveying it. And as a consequence, it is harder for the speaker to retract from conveying it without suggesting bad faith (unless she concedes that she did not master the intuitive procedure of inference in communication, which is self-face-threatening, but actually happens occasionally to everyone in ordinary conversations, and is simply a misunderstanding).

Yet the legitimacy of an intuitive feeling of bad faith is unlikely to be proved, and furthermore it is certainly more or less present to the consciousness of the speaker, who may sometimes be completely aware of her behaviour but sometimes speaks with a lower level of self-monitoring (more 'intuitively' in the usual sense of

that term). Even in the case above, there is no objective criterion that could be used to formally decide between a conscious attempt at bringing forward considerations about the costs of the wedding, which is a potential harm to the father-son relationship, a vaguely intuitive behaviour awkwardly exhibiting a worry to some lower degree of consciousness, and a comment that could be just intended as small talk without weighing the potential inferences in full. Retractions that appear commonsensically as bad faith are not *necessarily* such.

Sometimes, explicitly preventing unwanted but likely inferences can also trigger feelings of bad faith: it is funny to point out that when speakers utter "nothing implied", they make manifest that undesirable pragmatic meanings are actually invited by the sentence in the circumstances, but that they don't want to commit themselves to them—and it is often a way to actually imply without taking formal responsibility for it and at the same time making the whole process fully manifest. And, clearly, things like "nothing implied" sometimes trigger feelings of bad faith that the speaker even endorses to be obviously insincere—as long as it cannot be proven.

Therefore a retraction may be spotted as insincere and still allow preventing an accusation of lying, for example.

Constraints on retractions are stronger with explicatures than with implicatures, since we are closer to what was verbalized (explicatures, let us recall, are (mostly) extensions of the original sentence). Suppose you meet John and you talk about Mary, whom you just recently met. He tells you: "Mary has four children". Then some days later you discover that Mary has five children and you meet John again, and complain that he lied to you. If John is a logician, he could tell you that the pragmatic meaning was cancellable, for the reason that having x children does not entail not having x + n children. But unless he is a complete failure in terms of theory of mind and metarepresentation, he will not be able to claim in all fairness that he never intended to mean that Mary had four and only four children, without being suspected of blatant bad faith. It is not that John cannot pretend that he did not *mean* something, but rather that he can hardly pretend unawareness of the pragmatic meaning; he cannot pretend that he did not predict that the meaning would be spontaneously derived and therefore he cannot assume that he is not committed to that particular meaning. The process is globally the same as with implicatures except that with explicatures, there is very little room for doubts about what is meant in the context, for the reason that they are mere add-ons to what is formally presented (there is no deductive inference going on, just the grasping of a more precise meaning on the basis of a general proposition). The speaker is thus highly committed to explicatures, and he cannot assume that others will not assume this.

There is of course a whole range of cases where retractions trigger no feeling of bad faith at all (proper misunderstandings, implicatures that compete with other possible ones in the circumstances…) or feelings of bad faith so strong that they can actually be converted into accusations of lying, as in the case of *Mary has four children*. The case of the father complaining about the wedding is somewhere in between on the scale. In any case, the feeling of bad faith has something to do with

an impression of a discrepancy between the relevance of some particular inference and the difficulty to find other ways to obtain relevance.

De Saussure and Oswald (2009) suggest that while wrong cancellations—which obviously hardly happen at all in real speech situations—trigger formal inconsistencies, unfair retractions trigger a pragmatic form of inconsistency.

Consider example (3). Here, *many* is interpreted by the author of the SMF as *most*. It is a typical pragmatic inference whereby the first speaker resorts to a common type of understatement, that is, she tries to convey a stronger quantification than explicitly stated. There are many scalar pragmatic meanings based on quantification, and the way they are enriched can be to the *more* or to the *negation of the more* depending on all sorts of information in the context. An utterance like *Some students of mine are damn good* can be interpreted either as meaning *Not all my students are good* (the typical 'scalar implicature' of the pragmatic literature) or as an understatement, for *Most of my students are damn good*. Both of these pragmatic meanings can be cancelled, and retracted to various degrees given the expectations of relevance in the circumstances.

The interesting thing is that if someone uses such interpretations to form a SMF as in example (3), retraction on the part of the original speaker seems a truly hard work, a 'burden'. A possible explanation for this would be that the author of the SMF displays that he took *many* to be an understatement for *most*, which could be legitimate in a number of contexts. However, interestingly enough, the intuition suggests that his taking *many* for *most* is actually *not* merely a misunderstanding. Rather, it appears to the audience as a possible interpretation that was not appropriately controlled (i.e. prevented) by the speaker. In sum, the author of the SMF shows that the speaker did not foresee one possible interpretation of the utterance, as Aikin and Casey (2011) note, pointing out that this is an *ad hominem* component of the SMF.

All in all, things look like this: the author of the SMF suggests, by this pragmatic enrichment, that the speaker uttered *many*, without intending to make it a decipherable understatement for *most*, but actually thinks *most*. The SMF therefore exhibits that the speaker awkwardly tried to cover some deeper, real, thought of hers, which is about *most*, by means of using a quantifier perhaps more acceptable in the circumstances, namely *many*. Therefore the SMF bears an accusation of attempting to manipulate the audience, and at the same times suggests that the speaker performed badly in doing so. Therefore the SMF portrays the targeted speaker as both malevolent (she tries to hide a thought which is relevant to the discussion) and pragmatically incompetent.

8.3.2 Targeting the Individual

Dennett (1989) claims that interpreters simulate the behaviour of speakers as a strategy to attribute beliefs and intentions to them (the 'intentional stance'), in order to adopt an adequate interpreting strategy. Sperber (1994) recalls that if hearers

follow paths of least effort, in ideal circumstances, they should adopt a strategy which is both *naïve*, i.e. they assume that the speaker is benevolent, and *optimistic*, i.e. that the speaker is competent on the topic. We follow Padilla Cruz (2012) in suggesting that according to a constant monitoring of believability, which Sperber et al. (2010) call 'epistemic vigilance', hearers might adapt their interpretive strategy and switch from *naïve* to *cautious* (when raising doubts about the speaker's benevolence) and from *optimistic* to *pessimistic* (when raising doubts about the speaker's competence) or—of course—a combination of both.

At the same time, speakers continuously attempt to appear at least benevolent, and if they want to be convincing or persuasive, they will also attempt to appear competent. Let us shortly elaborate on this important aspect of argumentative interactions.

Trying to be convincing or persuasive is probably the most crucial property of what makes a context 'argumentative'. In such contexts, of course, roles alternate: the speaker's perlocutionary aims are to convince and persuade, while the hearer processes the utterance with epistemic vigilance. Doing so, not only does he infer, evaluate, accept, reject, etc. arguments, but he also tracks various signs, external to the content of the arguments themselves, in order to get feedback on the speaker's reliability, which is basically about assessing his *benevolence* and *competence*. Mirroring this, the speaker intuitively sends such signals and attempts to monitor their production.[12]

In short, in argumentative contexts, the speaker and the hearer, each in their turn, enter in a competition:

(a) The speaker who prefers the hearer to adopt a naïve optimistic strategy of interpretation and thus lower his epistemic vigilance; and
(b) The hearer who intuitively tracks signs of the contrary in order to best adapt his attitude of epistemic vigilance.

Not being competent is a problem of course but it can be avowed and assumed to a large extent, as it does not bear on moral principles. On the contrary, malevolence cannot. Therefore, an accusation bearing on benevolence is stronger in damaging the credibility of the speaker than one of incompetence (hence the success of *ad personam* fallacies). If the SMF targets an individual as we suggested, then it has the consequence of damaging the speaker's credibility on both levels: malevolence (the speaker tried to hide some less acceptable thought) and incompetence, but a specific type of incompetence: pragmatic incompetence, i.e. a cognitive type of incompetence leading to wrong predictions on what an interpreter can make of her utterance.

The irony of it is that the SMF is a fallacious accusation of fallaciousness; it is a non-cooperative move performed to allege non-cooperation. Since the SMF portrays the target as having low interpretative, cognitive, skills, and since it does so by uncovering a hidden meaning or thought, it displays by contrast the high pragmatic skills of its author.

[12]On the notions discussed in this paragraph, see Sperber et al. (2010).

Escaping an accusation of being tricky and manipulative under such circumstances is actually a burden. The *burden of proof* effect with the SMF is therefore not residing in the fact that the speaker merely has to correct an interpretation but that she has to resist an accusation of being misleading. As a consequence, she needs not only to correct the interpretation (she does so, of course), but furthermore to show that her opponent uses a wrong interpretation with the aim of unfairly accusing him of hiding thoughts that would be actually relevant in the current discussion.

Reacting to such an accusation requires, on the part of the original speaker, a switch from the ongoing argumentative interaction to the level of meta-argumentation: she must stop the current discussion in order to introduce a new topic, which is meta-discursive, and which is not about what is being discussed, but about how it is discussed, whether it is fair or not, etc.

This is very similar to what happens with fallacious presupposition accommodation. Consider (7) in this respect:

(7) The problem with our egalitarian society is that it disempowers the people.

Here, there is a claim (the problem with our society is that disempowers the people) and a justification is presupposed (our society is egalitarian). If the interlocutor is agnostic about the justification, then the presupposition has chances of being incorporated in the hearer's background without cautious evaluation. But the interesting thing is that if the hearer has a view on this, or if he wants to question the presupposition, he needs to escape the normal flow of the dialogue in order to enter into a metadiscursive negotiation.[13] This has a serious cost for both participants of the dialogue: it threatens both interlocutors' faces, for two directly related reasons: first, it implies that the speaker made wrong assumptions about what belongs to the common ground, and at the same that the hearer should have something in his common ground which he actually does not have, and, second, the hearer is basically saying that the speaker's utterance poses problems of relevance and informativeness. There is, therefore, a 'face-oriented' *burden* also posed to anyone questioning fallacious presuppositions. This happens also with *discursive presuppositions* which are weak implicatures or background assumptions that are necessarily drawn from what is said in order not to obtain a meaning but to obtain relevance (for example, the notion that guns are forbidden in these premises activates a background assumption that they may be permitted elsewhere); see de Saussure (2013) for elaborations on this notion.

Yet it must be noted on the other hand that argumentative discussions involve naturally face-threatening acts; raising such acts in an argumentation, especially when being personally targeted by a SMF, which is itself a dramatic face-threatening act, should not be an issue in the first place and therefore should not be the cause of the 'burden'. We will come back to the issue of the burden of proof further down, but let us already observe that this is an important difference with wrong or unfair

[13]This is a problem well-known to pragmaticists since the works of Ducrot (1972, 1980) on presuppositions, and it has been discussed at large in the literature since then.

presupposition accommodations, which can (i) occur outside of harsh debates and (ii) be intertwined with in-group versus out-group problems (rejecting a presupposition may socially amount to rejecting a tacit agreement inside a group).

In sum: the author of a SMF appears as the witty person who is uncovering what the speaker has in mind and which she (malevolently) tries to cover or hide; the author of the SFM puts the thoughts of the speaker in full light, therefore ridiculing her. It does so by showing that the speaker has bad cognitive skills, while the author of the SMF appears as mastering the whole path of reasoning available, starting from an utterance in a context and including the recovery of (possibly imagined) contextual premises. As a consequence, the author of the SMF gains prestige in two related dimensions: uncovering the hidden, and being particularly apt in doing so. The SMF has then a secondary consequence: its author tries to come across as an excellent arguer.

From the prestige so gained, the author of the SMF achieves a social victory, regardless of the validity of the interpretation sustaining the SMF.

Strikingly, it may even be that the author of an SMF does *not* hold the content of the SMF (he probably very seldom holds it true) and nonetheless wins the argument simply because of the prestige gained by exerting and showing his pragmatic, cognitive, skills, which are underlying components of argumentative skills.

8.4 The Burden of Proof and the Lack of Relevance of Justifications

So far, we have tried to explore the main pragmatic effects of a SMF: it targets the pragmatic skills of the speaker, which has two consequences: the speaker can only react by a complicated act of switching to a meta-discursive discussion, and the author of the SMF gains prestige by displaying his pragmatic abilities, which are about (allegedly) uncovering hidden thoughts of the speaker and showing that the speaker was unable to control the availability of some pragmatic interpretations which are dangerous for her position.

This can be a sound explanation for the fact that the SMF is a winning move, but still, it does not account satisfactorily for the particular weight of the burden of proof.

A typical error of the victim of a SMF is to still assume that it will be overcome by explaining that the interpretation is undue, but this would correct only the minor dimension of the SMF and not its major effect, i.e. social prestige.

Strangely enough, countering this prestige is very difficult. This is surprising because after all, argumentative interactions between skilled arguers often show an alternation of winning moves and virtually each move can make the preceding winner be the next loser. After each move, another argument, fair or fallacious, can achieve this reversal. Good arguments are certainly ultimately better in achieving this aim (see Mercier and Sperber 2011), but good arguments are also, sometimes, more complex and difficult to unfold.

Therefore, that the SMF is a winning move in an argumentation should not, in principle, prevent the speaker to react appropriately and serve a new argument which will constitute a winning move in its turn, therefore re-establishing the argumentative equilibrium. Yet it seems precisely to prevent such a move, at least in an unexpected measure. There must be something more specific to the SMF (and certainly to some other fallacies as well) that specifically impairs this possibility, and which the literature calls the *burden of proof*.

The fact that the speaker has to switch the level of the discussion from the topic being discussed to the way it is discussed, i.e. that she has to operate a move from discussing P to discussing the discussion about P, is of course the major cause of this burden, but we suggest that the SMF makes it *particularly difficult* for one reason, which is that the original speaker has to make *justifications* about what she said and what she meant, which are not only impossible to prove, but which will look overall irrelevant—therefore miss the objective of regaining prestige in front of an audience. Only *making points* is relevant 'in its own right'.

In terms of expectations, *making a point* is a conversational contribution in the full sense, raising expectations of relevance of its own. Making a point means updating the context with something new, adding it to the conversational background, i.e. to the commitment store (in Hamblin's 1970 terms).

Producing a SMF amounts to making a very strong point, explicitly about the thought that is attributed to the original speaker, and implicitly about her inability to predict that this thought could be derived from her verbal behaviour. On the contrary, reacting by embarking into a justification about why the interpretation presented by the author of the SMF is unfair, undue, incorrect or false is not 'making a point'. It is an operation of metadiscursive justification. And precisely, making a justification is by nature—because of its logical discursive structure—an argumentative move that is subordinate to some other claim (or 'point'). Making a justification is an elaboration of a point, not a point. In fact, if the speaker elaborates on why the interpretation is incorrect, she is merely commenting on the point, rather than making a point herself.

Still by nature, making a comment triggers less attention than making a new point.[14] In other terms, it is much more difficult for someone to raise expectations of relevance when *explaining the reasons why P was asserted* than when *asserting P*. So to say, the effort one would spend in processing the complexity inherent to justifications, in particular when one has to endure a stop in the flow of the interaction in order to pay attention to a metadiscursive type of justification, is not expected to trigger in return enough rewards in terms of knowledge for one's cognitive environment.

A speaker who begins to justify why her words were not correctly interpreted encounters a problem of attention also because it is common knowledge that it is virtually beyond any proof. It is very difficult, if not impossible, to prove that one

[14]Certainly, a justification is a point in some way, but since it is directly subordinate to some other point, we will not call them *points* proper.

does not believe Q. Trying to prove that Q does not follow from P, which she uttered, is a complicated thing even for ordinary, not fallacious (at least intentionally) implicatures. In ordinary conversation, you merely say "That's not what I meant"; but it is unlikely that a long explanation about *why the interlocutor should not have derived the implicature* is going to literally prove anything.

With a SMF, since the move was not benevolent, the escape cannot simply be to say that it was not meant. Trying to explain that the conclusions held by the opponent are not in one's head is obviously going to be a total failure.

In practice, what could be a successful countering of a SMF?

Given the picture we have delineated above, the answer lies in targeting the skills, not directly the contents. Irony and rhetorical questions seem good candidates, even though their success is difficult to guarantee. In order to react from a wrong interpretation of quantifiers, as in (3), the speaker might perhaps say something like "You are good at maths, aren't you". In front of a SMF, asking a question like "Waow, did I actually say *this*? Amazing!" ironically implying that the author of the SMF is telling you what you are supposed to mean, might have some potential to ridicule him in turn and reverse the argumentative equilibrium or even put the author of the SMF in greater trouble. However the risk of escalating the metadiscursive argument is real and may have problematic consequences of turning into a series of mutual accusations which would be detrimental to both arguers if the setting is one of talking in front of an audience whose agreement is sought and who may then view the two arguers as playing childish games.

8.5 Conclusions and Cognitive Perspectives

Going back now to the main cognitive question raised by fallacies and in particular by the SMF, which is to understand why humans, who have such cognitive abilities, can be prone to accepting, or surrendering, to fallacious argumentative moves, a number of explanations, although rather general, are on offer.

Certainly, the existence of fast and frugal heuristics (Gigerenzer 2004), and more broadly the general economy of cognitive processes, are key parts of the explanation. The processing of arguments, and in particular of manipulative ones and of fallacies, is starting to be addressed with new cognitive approaches such as Maillat and Oswald's (2009, 2011) notion of *context selection constraint*,[15] which shows how various strategies can be put in place by a speaker to prevent the audience from accessing propositions that would actually be relevant to them in order to make wise rational decisions about the acceptability of what the speaker says or manifests. Our account falls nicely with Dessalles' (2011) approach for which arguments serve primarily to track manipulative attempts and gain social benefits out of it, but also with Sperber et al.'s (2010) and Mercier and Sperber's (2011) approach; they

[15]See also de Saussure (2005).

develop a theory (the *Argumentative Theory of Reasoning*), with impressive experimental support, which postulates that the human ability to reason evolved as a means to argue with others, which entails an advantage in attaining truths via collective reasoning. Their approach convincingly predicts that group reasoning works better than individual reasoning, and, at the same time, that the outcome of argumentation is achieved better by sound, rather than bad, arguments. Yet, this approach also predicts the existence of cognitive biases, for the reason that the achievement of argumentation is actually better obtained in the presence of biases. Mercier explains:

> If reasoning evolved so we can argue with others, then we should be biased in our search for arguments. In a discussion, I have little use for arguments that support your point of view or that rebut mine. Accordingly, *reasoning should display a confirmation bias*: it should be more likely to find arguments that support our point of view or rebut those that we oppose. Short (but emphatic) answer: it does, and very much so. The confirmation bias is one of the most robust and prevalent biases in reasoning."[16]

They emphasize that an individual has little interest in arguments that do not support one's own standpoints.

It could be that the SMF, if we are right that it does not actually target arguments themselves but pragmatic skills, is a counterargument to Mercier and Sperber, but in fact it is not. In fact, the SMF is not powerful at *convincing* others in an argumentative interaction. Rather it is successful in order to have the opponent surrender in front of an audience for causes that are alien to the validity of the arguments themselves. The SMF, we claim, manifests a superior skill in manipulating arguments on the part of its author, which is the basic cause of the opponent's defeat in the eyes of an external audience. This makes clear sense within the Mercier and Sperber view, since pragmatic skills are valuable in themselves as a warrant for good argumentative skills, despite the fact that in some cases the argument itself is not successful in actually persuading (sometimes, of course, a SMF is successful at doing that).[17]

In this paper, we argue that the SMF is a 'pragmatic winner' not because of its potential persuasive force, but rather because it targets the pragmatic cognitive-inferential skills of its victim while enhancing the prestige of its author.

In particular, we claim that:

(i) the burden of proof has essentially to do with the difficulty to get the attention of the audience back, because
(ii) a reaction to the SMF is mostly a discussion of the reasons why the inference drawn was unduly derived,

[16]'The argumentative theory of reasoning blog': https://sites.google.com/site/hugomercier/theargumentativetheoryofreasoning.

[17]Both *better outcomes in collective decision making* and *targeting manipulative attempts* are operations that go hand-in-hand in argumentation; that one is the main evolutionary reason of argumentation is out of the scope of this paper, so we will not attempt at discussing the various positions on offer here in this respect.

(iii) which is virtually destined to be a failure first of all because of the lack of relevance of *justifications* in comparison with that of actual points, and because of the intrinsic difficulty of arguing about one's meaning intentions and thoughts on the basis of some behaviour.

These conclusions await, of course, further empirical validation.

We also suggested that the victim of a SMF should probably also favour a reaction targeting pragmatic and cognitive skills, such as irony or rhetorical questions.

One thing that we did not discuss is the actual format of inferences grounding SMFs. It appears that sometimes these inferences are drawn just like normal implicatures or explicatures, i.e., either by complementing the linguistic logical form with some contextual fine-tuning, as when "Some like it hot" is understood as conveying "not all persons like it hot". Many inferences are however not achieved through processes that correspond to logical patterns but through non-logical heuristics, such as analogies. All these inferential ways have emerged because they have positive outcomes: many fallacies lead to bad results in some contexts but are successful pieces of knowledge in others. This makes the identification of a SMF very tricky. Ultimately, the SMF relies on the ability of the audience to grasp the inferential path, logical or not, followed by the author of the SMF. The effects of connivance that are, as a result, also achieved by the SMF make it even more difficult to counter, because the victim finds himself ousted of the group and targeted as someone to be pessimistic and cautious about, while the author itself attracts the prestige of denunciation on top of the rest.

References

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

de Saussure, Louis. 2005. Manipulation and cognitive pragmatics: Preliminary hypotheses. In *Manipulation and ideologies in the twentieth century. Discourse, language, mind*, ed. Louis de Saussure & Peter Schulz, 113–146. Amsterdam, Philadelphia: John Benjamins.

de Saussure, Louis. 2013. Background relevance. *Journal of Pragmatics* 59 (Part B): 178–189. https://doi.org/10.1016/j.pragma.2013.08.009.

de Saussure, Louis, and Steve Oswald. 2009. Argumentation et engagement du locuteur. Pour un point de vue subjectiviste. *Nouveaux Cahiers de Linguistique Française* 29: 215–243.

Dennett, Daniel. 1989. *The intentional stance*. Cambridge: The MIT Press.

Dessalles, Jean-Louis. 2011. Reasoning as a lie detection device (Commentary on Mercier & Sperber: 'Why do humans reason? Arguments for an argumentative theory'). *Behavioral and Brain Sciences* 34 (2): 76–77.

Ducrot, Oswald. 1972. *Dire et ne pas dire*. Paris: Hermann.

Ducrot, Oswald. 1980. *Le dire et le dit*. Paris: Minuit.

Gigerenzer, Gerd. 2004. Fast and frugal heuristics: The tools of bounded rationality. In *Blackwell handbook of judgment and decision making*, ed. Derek J. Koehler, and Nigel Harvey, 62–88. Oxford: Blackwell.

Hamblin, Charles. 1970. *Fallacies*. London: Methuen.

Lewiński, Marcin and Steve Oswald. 2013. When and how do we deal with straw men? A normative and cognitive pragmatic account. *Journal of Pragmatics* 59(Part B): 164–177. https://doi.org/10.1016/j.pragma.2013.05.001.

Maillat, Didier, and Steve Oswald. 2009. Defining manipulative discourse: The pragmatics of cognitive illusions. *International Review of Pragmatics* 1 (2): 348–370.

Maillat, Didier, and Steve Oswald. 2011. Constraining context: A pragmatic account of cognitive manipulation. In *Critical discourse studies in context and cognition*, ed. Chris Hart, 65–80. Amsterdam: John Benjamins.

Mercier, Hugo, and Dan Sperber. 2011. Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences* 34: 57–111.

Morency, Patrick, Steve Oswald, and Louis de Saussure. 2008. Explicitness, implicitness and commitment attribution: A cognitive pragmatic account. *Belgian Journal of Linguistics* 22: 197–219.

Oswald, Steve, and Marcin Lewiński. 2014. Pragmatics, cognitive heuristics and the straw man fallacy. In *Rhetoric & Cognition: Theoretical perspectives and persuasive strategies*, ed. Thierry Herman, and Steve Oswald, 313–343. Bern: Peter Lang.

Padilla Cruz, Manuel. 2012. Epistemic vigilance, cautious optimism and sophisticated understanding. *Research in Language* 10 (4): 365–386.

Petty, Richard E., and John T. Cacioppo. 1986. The elaboration likelihood model of persuasion. *Advances in Experimental Social Psychology* 19: 123–205.

Pinker, Steven, Martin A. Nowak, and James J. Lee. 2008. The logic of indirect speech. *Proceedings of the National Academy of Sciences of the United States of America* 105 (3): 833–838. https://doi.org/10.1073/pnas.0707192105.

Reboul, Anne. 2017. *Cognition and communication in the evolution of language*. Oxford: Oxford University Press.

Sperber, Dan. 1994. Understanding verbal understanding. In *What is intelligence?*, ed. Jean Khalfa, 179–198. Cambridge: Cambridge University Press.

Sperber, Dan & Deirdre Wilson. 1995. *Relevance. Communication and cognition* (1st ed.: 1986). Oxford: Blackwell.

Sperber, Dan, Fabrice Clément, Christophe Heintz, Olivier Mascaro, Hugo Mercier, Gloria Origgi, and Deirdre Wilson. 2010. Epistemic vigilance. *Mind and Language* 25 (4): 359–393.

Tversky, Amos, and Daniel Kahneman. 1974. Judgment under uncertainty: Heuristics and biases. *Science* 185 (4157): 1124–1131.