

Steve Oswald  
Cognitive Science Centre, University of Neuchâtel  
&  
Christopher Hart  
Lancaster University

## Trust Based on Bias: Cognitive Constraints on Source-related Fallacies

### 0. Introduction

Much of the research concerned with fallacies has been typological in nature, attempting to delineate particular kinds of fallacies. Among the fallacies identified are those which relate to the reliability of third party sources, either by making appeal to those sources (*ad populum* and *ad verecundiam*) or attempting to discredit them (*ad hominem*). This research has been focused on specifying the characteristics of these arguments, including, for example, what it is that makes them fallacious (Walton, 2006; van Eemeren & Grootendorst, 2004). This paper will approach these fallacies – referred to hereafter as source-related fallacies – within an explanatory framework in which rhetorical effectiveness is seen as a product of cognitive constraints and biases (see Hart, 2011, forth.; Oswald, 2011, forth.). We will therefore not be concerned with descriptive and definitional issues. Rather, we aim to illuminate *how* and *why* illegitimate invocations of third parties may succeed in convincing the audience of a given conclusion.

This framework draws on insights from cognitive science, in particular cognitive pragmatics (Sperber & Wilson, 1995; Wilson & Sperber, 2012) and evolutionary approaches to communication (Sperber *et al.*, 2010). We will argue that there are cognitive principles pertaining to the use and processing of source-related fallacies which account for their rhetorical effect. These principles take the form of cognitive heuristics (see Gigerenzer, 2008) which guide inferential processes at the level of evaluation.<sup>1</sup> Specifically, we are suggesting that source-related fallacies exploit our processing mechanisms, which are usually reliable and accurate, but which are by their necessarily fast and frugal nature inherently fallible.

In section 1 we introduce the three fallacies we are concerned with. Section 2 presents an evolutionary approach to argumentation which suggests that *ad populum* and *ad*

---

<sup>1</sup> See Oswald (2011, forth.) on constraints operating at the level of comprehension.

*verecundiam* arguments meet the demands of hearers' systems for *epistemic vigilance* (Sperber *et al.*, 2010) by presenting sources mistakenly deemed reliable whilst *ad hominem* alerts epistemic vigilance to characteristics of the source targeted which misguidedly betray their unreliability. In section 3 we elaborate an account of fallacy processing in terms of cognitive constraints and biases. Here, we focus on the information-selection mechanisms at play in argument processing. We suggest that these fallacies work, as a consequence of evolved cognitive biases (see e.g. Pohl, 2004) and relevance considerations in processing, by constraining the information sets available to the hearer, resulting in a hindrance to critical evaluation. Section 4 illustrates our claims with an analysis of each argument as found in political discourse.

## 1. Source-related fallacies

Source-related fallacies may be identified as those whose premises invoke third parties in the course of the argument. These can be divided into arguments which, in order to support the speaker's conclusion, resort to third parties in the form of authorities and majorities (*ad verecundiam* and *ad populum* respectively) and those which attack third parties (*ad hominem*). Following Walton (2006), the fallaciousness of these arguments can be captured in their failure to satisfactorily answer critical questions associated with underlying argument schemes.

*Ad verecundiam*, for example, can be broadly construed as a problematic appeal to expert opinion. Of course, not all appeals to expert opinion are fallacious. What determines the fallaciousness of an appeal to expert opinion, according to Walton, is whether the it satisfactorily answers a set of critical questions associated with an underlying argumentation scheme. The argumentation scheme behind appeals to expert opinion can be presented as below (Walton, 2006, p. 87):

(1) "Argumentation scheme for appeal to expert opinion

Major premise: Source E is an expert in subject domain D containing Proposition A.

Minor premise: E asserts that proposition A (in domain D) is true (false).

Conclusion: A may plausibly be taken to be true (false)."

The critical questions that determine whether or not a particular instantiation of the scheme is fallacious are formulated as follows (Walton, 2006, p. 88):

(2) "Expertise Question. How credible is E as an expert source?

Field Question. Is E an expert in the field that A is in?

Opinion Question. What did E assert that implies A?

Trustworthiness Question. Is E personally reliable as a source?

Consistency Question. Is A consistent with what other experts assert?

Backup Evidence Question. Is E's assertion based on evidence?"

*Ad verecundiam* is the fallacious variant of the argument scheme which arises when any one of these questions receives a negative answer. Walton formulates similar schemes and associated critical questions to assess the fallaciousness of appeals to popular opinion and attacks of the source of a given proposition. For reasons of space, we do not present these here but refer the reader directly to Walton (2006). However, we note that *ad populum* and *ad hominem*, like *ad verecundiam*, are the fallacious variants of respective underlying argument schemes.

One critical question which crops up in all three argument schemes concerns the reliability and/or credibility of the third party. In all three, what is at stake is not the content of the argument but, rather, something external to it. In the case of *ad verecundiam*, the strength of the argument lays in the perceived trustworthiness of the source. The strength of *ad populum* arguments lays not in the faith one has in a particular source but in the weight carried by widespread belief. *Ad hominem* similarly operates on the parameter of authority but in the reverse direction where it dismisses a given proposition on the grounds that it is asserted by a source whose legitimacy is cast into doubt.

In the next section, we will argue that the critical abilities underlying Walton's critical questions are encapsulated in the evolved, cognitive competences of communicators. These competences may take the form of what Sperber *et al.* (2010) refer to as 'epistemic vigilance'. Source-related fallacies, we will suggest, work by manipulating epistemic vigilance where they can be characterised as "failed diagnostic strategies" (Jackson, 1996, p. 111) reinterpreted in cognitive and adaptationist terms.

## **2. Epistemic vigilance**

On an evolutionary account of communication, language must have evolved initially for purposes of cooperation (Sperber, 2001; Hurford, 2007). However, any such system of cooperation is susceptible to exploitation in the form of deception (Origg and Sperber, 2000). For communication to stabilise within the species, therefore, "all cost-effective available modes of defence are likely to have evolved" (Sperber, 2000, p. 135). For Sperber *et al.* (2010: 359) these defences amount to "a suite of cognitive mechanisms for epistemic vigilance". Epistemic vigilance, according to Sperber *et al.*, is both content-directed and source-directed. In content-directed vigilance, hearers attend to the logical consistency of the message as well as its plausibility given background assumptions. In source-directed vigilance, hearers assess the trustworthiness of speakers, defined as judgements of competence, benevolence, expertise, reliability, credibility, etc.

In the practice of argumentation – a communicative practice by which speakers provide information in support of claims in order to convince others – hearers exercise epistemic vigilance in order to avoid being misled. The functioning mechanism here is an argumentative module which operates both in the production and evaluation of

arguments. In the evaluation of arguments, the mechanism is responsible for identifying reasons to (dis)believe the conclusion of an argument (Mercier and Sperber, 2009, 2011). The module, here, takes as input claims and combines them with contextual information to yield an intuitive representation concerning the relationship between the premises and the conclusion contained within the argument. This then allows the hearer to make a judgement on the argument's acceptability. In the production of arguments, the argumentative module allows speakers to formulate arguments in ways which are intended to satisfy or otherwise exploit the hearer's epistemic vigilance.

In so far as we are interested in the rhetorical effectiveness of fallacies, we are interested here in the evaluative dimension of the argumentative module. Moreover, because we are concerned with source-related fallacies in particular, we will deal only with source-directed vigilance.

Epistemic vigilance amounts to a diagnostic strategy selected for in the evolution of communication. However, by virtue of the resource-bound nature of our cognitive architecture, it remains fallible (see section 3). It is precisely this imperfection, we suggest, that fallacies exploit. *Ad verecundiam* and *ad populum* appear to provide sufficient evidence to satisfy epistemic vigilance by presenting the conclusion as entertained by external sources the hearer is expected to trust. In *ad verecundiam*, it is the expertise, authority, competence, credibility and benevolence etc. of the source that lends epistemic strength to the conclusion. In *ad populum*, the epistemic strength lays, in part, in the sense of likelihood derived from learning that a large number of people take the conclusion to be true. *Ad hominem*, by contrast, directs hearers to the epistemic weakness of the conclusion advanced by the targeted source by making explicit their untrustworthiness. *Ad hominem* itself therefore satisfies epistemic vigilance directed at the speaker (the utterer of *ad hominem*) but alerts epistemic vigilance directed toward the targeted source.

The three fallacies, then, may overcome epistemic vigilance by providing apparently satisfactory evidence; the first two rely on perceived trustworthiness whilst the latter directs epistemic vigilance towards 'mistakenly' assessing the unreliability of the source as a sufficient reason to reject the conclusion they advance. When the fallacies are successful, then, it means that the cognitive system has not found it relevant to question the evidence provided within the argument. In other words, the system has failed to mobilise additional (critical) information which should alert the hearer to the argument's fallaciousness. To account for why hearers fail to recognise these arguments as fallacious, we turn to particular cognitive constraints and biases involved in argument processing.

### 3. Processing constraints and biases

One major constraint on argument processing consists in information selection (Oswald 2011, forth.; Maillat & Oswald, 2009, 2011). In this section we suggest that information selection may be influenced by evolved cognitive biases and considerations of relevance such that the importance of critical questioning is downplayed. As a general theory of human cognition, applied to communication, and to the extent that it formulates precise criteria of information selection, Relevance Theory (henceforth RT) allows a handle on argumentative phenomena of the kind we are concerned with here.

#### 3.1. Relevance Theory

RT is a model of ostensive-inferential communication which postulates that the cognitive mechanisms involved in comprehension are regulated according to *relevance*, defined as an optimal balance between cognitive costs and benefits. Understanding a speaker's utterance amounts to identifying the interpretation that best satisfies the ratio between the cognitive efforts incurred in processing the utterance (including the selection of contextual information required to make sense of it) and the cognitive effects that can be anticipated. Cognitive effects are defined in terms of an assumption's usefulness to the cognitive system (i.e., its ability to provide reliable new information and to revise, discard or strengthen previously held information). In interpretation, hearer's follow a *path of least effort* in calculating cognitive effects and cease further processing as soon as expectations of relevance are met.

This technical definition of relevance highlights two crucial parameters according to which information is selected. These are captured by the *extent conditions* of relevance (Sperber & Wilson, 1995, p. 125):

- (3) "Extent condition 1: an assumption is relevant in a context to the extent that its contextual effects in this context are large.  
Extent condition 2: an assumption is relevant in a context to the extent that the effort required to process it in this context is small."

Relevance, thus, is a function of accessibility, since that which is accessible is less effortful to process. At the same time, it is a function of epistemic strength, where adding, revising or discarding assumptions is, of course, an epistemic matter. Two important points emerge for our purposes:

- relevance assessments are heuristic in nature and therefore susceptible to error;
- the extent conditions of relevance may extend beyond the comprehension procedure to provide parameters for information selection in argument evaluation.

In relation to the second point above, we suggest that fallacies work precisely by prohibiting hearers from selecting critical information which would point to their fallacious nature. This argument is founded on the principle that information processing

involves competition between information sets. As Sperber and Wilson put it, “not all chunks of information are equally accessible at any given time” (1995, p. 138). Certain information is selected, according to its relevance in a particular context, at the expense of other information which, by the procedure, is inverse-proportionately less relevant and therefore not attended to. In what follows, we argue that this selection effect is, at least in part, a function of evolved cognitive biases.

### 3.2. Cognitive biases

Cognitive biases are errors in judgement, thinking and memory which can be predicted to arise from heuristics in information processing (Tversky and Kahneman, 1974; Pohl, 2004; Gigerenzer, 2008). Heuristics are decision-making *strategies* which are generally useful and have evolved to cope with uncertainty, resource limitations and time pressures etc. However, by their ‘fast and frugal’ nature, they can sometimes lead to “severe and systematic errors” (Tversky and Kahneman, 1974, p. 1124).

Certain cognitive biases bring about errors specifically in epistemic judgements. For example, a *conformity bias* can be shown to result in something like a ‘group effect’ (Henrich and Boyd, 1998). Such heuristics and biases may be rooted in the evolution of social cooperation. For example, as Sperber *et al.* (2010, p. 380) state:

“If an idea is generally accepted by the people you interact with, isn’t this a good reason to accept it too? It may be a modest and prudent policy to go along with the people one interacts with, and to accept the ideas they accept. Anything else may compromise one’s cultural competence and social acceptability.”

The claim we are making in this paper is that cognitive biases of this kind, in conjunction with relevance considerations in information processing, may explain the effectiveness of certain fallacies, including source-related fallacies.

Cognitive biases are triggered by fallacies and in turn provide additional input to the argumentative module. Given what is at stake, biases have evolved to make this input highly relevant to the argumentative module. As a result, in evaluating the argument the argumentative module may deem it unnecessary to engage in further cognitive processing, including asking critical questions. The argument satisfies expectations of relevance and, by consequence, epistemic vigilance too; it therefore goes through the system unchallenged.

On this account, fallacies are directly related to cognitive biases which make some information relevant and other, critical information irrelevant. *Ad populum* operates on the back of a conformity bias.<sup>2</sup> We can make a similar case for *ad verecundiam*. For example, the famous Milgram experiments (1974) observed an ‘obedience effect’ which is presumably, at least partly, the product of an authority bias. *Ad verecundiam* is able to

---

<sup>2</sup> Maillat (forth.) has suggested that *ad populum* also hinges on the bias behind the *validity effect* (Hackett-Renner, 2004) where repetition can be seen to contribute to acceptance of information.

activate precisely this bias. It usually pays to go along with majorities because to do so is efficient and because not to do so compromises one's social inclusion. It may pay to go along with authorities for similar reasons. Moreover, experts possess information in particular domains which hearers might not otherwise have access to and which, by virtue of their expertise, is usually reliable. It is therefore perfectly rationale to take the word of an expert. Fallacies, in sum then, work in convincing audiences because they exploit the inherent fallibility of heuristics which are, in the normal course of events, helpful.

Biases related to conformity and expertise make what majorities and authorities say or think epistemically strong to the degree that hearers are less likely to subject the argument to full critical evaluation. *Ad hominem* can be analysed as operating on the same mechanisms, though it may be considered the flipside of *ad verecundiam*. Whilst *ad verecundiam* tries to establish the source's trustworthiness, *ad hominem* tries to undermine their trustworthiness. And since the risks of communication are so great, any suggestion that the source is not credible is relevant enough for the system that the hearer does not properly consider the content of the argument or critically question the significance of the source's credibility in relation to the quality of the argument.

In the final section, we see how all of this plays out in particular illustrative analyses.

## 4. Illustration and analysis

### 4.1. *Ad verecundiam*<sup>3</sup>

In 2004, prior to a national vote about the possibility for 2nd and 3rd generation immigrants to benefit from simplified naturalisation procedures, the Swiss far-right political party UDC published a campaign ad in several newspapers entitled "Thanks to automatic naturalisation, the Muslims soon to be in majority?" ("*Grâce aux naturalisations automatiques, les Musulmans bientôt en majorité?*"). The document contains three graphs with statistics and a text explaining to the reader that simplifying these naturalisation procedures will result in a dramatic increase of Muslims in the country, which is something they implicitly present as an undesirable threat to the Swiss identity. In explicit terms, they merely refer to this as "a problem". Interestingly, the text mentions the testimony of an expert, Sami Aldeeb, who works for the Swiss Institute of Comparative Law at the University of Lausanne. Here is the text:

- (4) "Sami Aldeeb, who is responsible for Arabic and Muslim law at the Swiss Institute of Comparative Law in Lausanne, has an even more drastic vision of the situation: "The proportion of Muslims in Switzerland triples every ten years. Today, 310'000 Muslims legally reside in the country and some 150'000

---

<sup>3</sup> The example is taken from Burger *et al.* (2006), who perform a Speech act-theoretic analysis of the corpus. We reinterpret here some of their findings within our framework.

others illegally reside in Switzerland. In twenty years, they will be in majority. There will be more Muslims than Christians in Switzerland by then'. And that is where the problem lies, because 'Muslims place their religion above our laws.'"<sup>4</sup>

The argument contained in this excerpt is quite complex but minimally offers the premise in (5) as a main reason to defend the conclusion in (6):

- (5) Explicit premise: Muslims place their religion above our laws.
- (6) Conclusion: It is a problem that in twenty years Muslims will outnumber Christians in Switzerland.

The argument contained in (5) is part of a more global pragmatic argumentation developed throughout the text which enjoins readers to vote against facilitated naturalisations in order to avoid the undesirability (in the opinion of the authors) of a majority of Muslims. We are specifically interested here in (i) the perceived authoritative character of the expert that is called in and (ii) the way the argument links the premises to its conclusion.

Sami Aldeeb is presented as an expert in comparative law, which indeed he was at that time. His testimony on Muslim law should thus at first sight be perceived as relevant and trustworthy. As a consequence, the contents expressed between quotation marks are likely to be perceived as epistemically strong. What the reader doesn't know is that his words are actually quoted from a prior interview he gave to the Swiss tabloid *Blick*, the publication of which Aldeeb had not approved himself, because instead of reflecting his actual point, which consisted in the defence of an integrationist model, his words were used for their potential islamophobic connotations (Burger *et al.*, 2006, p. 17). We note that unless the reader personally knows Aldeeb, they have no way of knowing this and they therefore have no cause to address the critical questions associated with the relevant argument scheme. In other words, epistemic vigilance is satisfied. This is problematic, however, because the reader has been prevented from critically assessing the expert's testimony on the issue.

A second problem with the use of the expert source in this argument, as noted by Burger *et al.* (*ibid.*), lays in the tentative inferences the reader is likely to draw from the Arabic resonance of the name Sami Aldeeb. In the absence of further evidence, the reader might infer from his name that the expert is himself Muslim, which would therefore tend to represent him as someone who does not hold hostile preconceptions about Muslims, i.e., as a speaker who is unbiased towards Muslims. Under this interpretation, his words, which negatively represent Muslims, cannot be taken to be motivated by an inherent

---

<sup>4</sup>"Sami Aldeeb, responsable du droit arabe et musulman à l'Institut suisse de droit comparé à Lausanne, a une vision encore plus drastique de la situation: 'La proportion de Musulmans en Suisse triple tous les dix ans. Aujourd'hui, 310'000 Musulmans vivent légalement et quelque 150'000 autres vivent illégalement en Suisse. Dans vingt ans, ils seront la majorité. Il y aura alors plus de Musulmans que de Chrétiens en Suisse'. Et c'est bien là que réside le problème, car 'Les Musulmans placent leur religion au-dessus de nos lois.'"



dislike for Muslims. This, in turn, positively affects the objective quality of the expert and increases his perceived trustworthiness. Nowhere in the document is the reader discouraged from drawing this inference, which is actually false; Sami Aldeeb is in fact a Christian Palestinian. Again, this piece of information would be highly relevant for the reader to establish source credibility. Its absence only increases the chances of the false inference being drawn and as a consequence boosts the epistemic status of whatever the 'unbiased' expert has to say.

A third problem with this argument can be identified in the problematic endorsement of the argumentative link presented in (5). A close look at the text shows that Sami Aldeeb only provides contents which are used in the premises of the argument (the propositions given in quotation marks). The core component of the conclusion, including the argumentative connective 'because', is however not inserted in quotation marks. We hypothesise that inattentive readers might overlook this and take Aldeeb to be responsible for the argument being made for the following three reasons:

- his testimony is framed as a "drastic vision of the situation". Drastic and dramatic visions of a situation are likely to foreground problematic issues
- the negative connotations of the statements he is responsible for appear in bold characters, making them more salient than both the conclusion and the connective
- assuming the reader's epistemic vigilance filters are satisfied of Aldeeb's trustworthiness in the way described above, the credibility of his words and their islamophobic connotations is compatible with the interpretation that they constitute a serious problem

(4) is a particularly interesting illustration of the *ad verecundiam* fallacy. The features of the text highlight the trustworthiness of the source in a way that should meet the demands of epistemic vigilance.

#### **4.2. *Ad populum***

In 2005, Michael Howard, former leader of the British Conservative Party, delivered a speech on immigration in Telford in which he argued for a restrictive policy on immigration and specifically on the asylum system. The main standpoint of his discourse is found in the following excerpt:<sup>5</sup>

- (7) "It's not racist to talk about immigration. It's not racist to criticise the system. It's not racist to want to limit the numbers"

In order to support these claims, he verbalises the voice of the majority in the following ways:

- (8) "The majority of British people (...) are united on this issue."

---

<sup>5</sup> The complete transcript can be found here (last accessed 08.03.2013):  
[http://news.bbc.co.uk/2/hi/uk\\_news/politics/vote\\_2005/frontpage/4430453.stm](http://news.bbc.co.uk/2/hi/uk_news/politics/vote_2005/frontpage/4430453.stm)

- (9) “Everyone wants new people to settle as long as numbers are limited.”
- (10) “Talk to people and whatever their background, religion or the colour of their skin – they ask the same thing: ‘Why can't we get a grip on immigration?’ These are the people who are always ready to welcome genuine refugees to Britain, or families who want to work hard and make a positive contribution to our country.”
- (11) “But I’ve lost count of the times that British people of ethnic backgrounds have told me that firm but fair immigration controls are essential for good relations.”

The implicit overall argumentation produced by Howard in this speech is an appeal to popular opinion which can be minimally reconstructed in the following way:

- (12) Everyone, like us, says there is a problem with immigration: this cannot be racism, since it is unlikely that all of the people denoted by ‘everyone’ are wrong and racist.

We take this argumentation to be an instance of the *ad populum* fallacy for two reasons. First, it fails to satisfactorily answer the critical questions associated with the scheme for arguments from popular opinion (in particular, we do not have conclusive evidence that the large majority invoked by Howard actually endorses what he says). Second, the way popular opinion is portrayed by Howard systematically and repeatedly presents “everyone” as morally respectable. However, the grounds for this appraisal are ill-evidenced. Nevertheless, the moral qualities of the majority referred to confer epistemic strength on the arguments in the following ways.

The third party invoked in support of Howard’s conclusions is a majority depicted as: (i) morally virtuous, i.e., hard-working and welcoming (10); and (ii) non-racist and unbiased towards foreigners, because they are themselves of different ethnic backgrounds and origins (10,11). We have seen previously (see section 3) that the voice of majorities carries epistemic weight, to the extent that it is perfectly rational to consider that a large amount of people united on a given issue are unlikely to be wrong. Depicting a majority as morally virtuous further adds to the perceived trustworthiness of the source. This in turn ensures that the assumptions reported to be endorsed by the majority are highly salient in the hearer’s cognitive environment as he processes Howard’s speech.<sup>6</sup>

The cognitive bias responsible for the *validity effect* can here be seen to operate too: Howard repeats on several occasions in this discourse the idea that the majority of British people agree with him, and this could also be seen as a way of increasing the epistemic value of the majority’s opinion on the subject. Additionally, since repetition makes for salience and accessibility, the fact that Howard’s (and the majority’s) opinion is repeated across the speech makes it more accessible than alternative information sets.

---

<sup>6</sup> The cognitive environment is defined as the set of assumptions that are manifest to an individual at any given time (Sperber & Wilson, 1995).

From the perspective of cognitive processing, this means that it stands a better chance of being represented than other, less salient, information sets, which could contain critical information pertaining to the argument.

Finally, as mentioned above in section 3.2., the conformity bias is exploited in Howard's words in his positive characterisation of the majority whose opinion he reports. The social consequences of disagreeing with a morally good, unbiased and non-racist majority are potentially disastrous for an individual. In the Telford speech, Howard's depiction of the majority he summons in support of his claim could very well be seen as a way of constraining any potential critical reaction on behalf of the audience.

### **4.3. *Ad hominem***

During Prime Minister's Questions on October 18 2011, Russell Brown, Labour MP, asked David Cameron, Britain's current Prime Minister, if he could provide a list of people affected by the Adam Werritty scandal.<sup>7</sup> To this, Cameron replied with the following:<sup>8</sup>

(13) "It comes slightly ill from a party to lecture us on lobbying when we now know the former defence secretary is working for a helicopter company, the former home secretary is working for a security firm, Lord Mandelson, well he's at Lazards, and even the former leader, the former prime minister, in the last few months he's got £120,000 for speeches to Credit Suisse, Visa and Citibank. He told us he'd put the money into the banks, we didn't know he'd get it out so quickly."

Russell's question required the PM to provide information about some lobbying practices his party had been accused of. Instead, the PM evades the attack by highlighting the opposition's own dubious practices, and therefore their non-legitimacy in filing attacks on the issue. Cameron, then, has used the *tu quoque* variant of the *ad hominem* fallacy. This argument is fallacious because "the hypocrisy of the arguer is not necessarily evidence of the falsity of what she argues" (see Aikin 2008, p. 156).

By pointing to the hypocrisy of his attacker, Cameron is attempting to undermine their credibility. This is quite common in political discourse (and political debates in particular). However, (13) is remarkable for a specific reason. Not only does Cameron provide evidence of dubious practices of the opposition, he provides a *significantly higher number* of such practices in order to outweigh Russell's attack. In the situation where epistemic vigilance is required to assess the trustworthiness of third parties which are being charged with inconsistency or hypocrisy, one could reasonably assume that the most trusted source would be the one to suffer the least damage. By identifying

---

<sup>7</sup> Adam Werritty is a business investor who was a close friend to Liam Fox, the then Secretary of State for Defence. Werritty allegedly accompanied Fox on overseas engagements acting as an informal adviser. Werritty was able use his position to provide access to the minister. This conflict of interest prompted a series of investigations which lead eventually to Fox's resignation in October 2011.

<sup>8</sup> See the transcript of the argument here (last accessed 08.03.2013): <http://www.heraldsotland.com/politics/political-news/cameron-attacks-brown-on-speeches.15458313>

numerous members of the opposition and the dubious practices they are accused of, Cameron overwhelmingly outcharges his opponent. (13) further sets the stage for a generalising inference to be drawn where hearers might extend the negative traits of some members of the opposition to the whole party.

Cognitively speaking, then, (13) foregrounds information meant to destroy the opponent's legitimacy. The *tu quoque* attack is particularly strong here from an epistemic perspective because it presents different instances of the same wrongdoing, whilst the original attack it addresses only questions the extent to which the Werritty scandal has damaged the ruling party. For reasons of relevance related to the possibility and usefulness of identifying concrete instances of the problem, this may prevent the audience from mobilising critical information. They might, accordingly, be left with no grounds to critically question the argument produced by Cameron. This, in turn, would result in the argument not being challenged and is precisely what makes the fallacy effective.

## 5. Conclusion

According to Hamblin, the standard treatment of fallacies is to suggest that a "fallacious argument ... is one that *seems to be valid* but *is not so* (Hamblin 1970, p.12, original emphasis). This, he argues, does not allow for the elaboration of a consistent theory of fallacies. However, we take his statement to highlight a crucial feature of fallacies which precisely explains their effectiveness: hearers fall for fallacies because they do not spot them. This can be accounted for by examining the evolved, cognitive constraints involved in information processing.

The heuristics and biases programme and RT together provide an explanatory framework for interpreting the rhetorical effectiveness of fallacies which we characterise in terms of the consequences they have for epistemic vigilance. We have considered the case of source-related fallacies and argued that these implement constraints on the information sets that will eventually be deemed relevant as hearers process the argument. These constraints entail that critical questions, which we see as embodied in epistemic vigilance, are neither asked nor answered.

## References

- Aikin, S. (2008). Tu Quoque Arguments and the Significance of Hypocrisy. *Informal Logic*, 28 (2), 155-169.
- Burger, M., G. Lugrin, R. Micheli & S. Pahud (2006). Marques linguistiques et manipulation. Le cas d'une campagne de l'extrême droite suisse. *Mots. Les Langages du Politique*, 81, 9-22.

- Eemeren, F. van and Grootendorst, R. (2004). *A Systematic Theory of Argumentation. The Pragmadiialectical Approach*. Cambridge: Cambridge University Press.
- Gigerenzer, G. (2008). Why heuristics work. *Perspectives on Psychological Science*, 3 (1), 20-29.
- Hackett Renner, C. (2004). Validity effect. In R. F. Pohl (Ed.), *Cognitive Illusions: A Handbook on Fallacies and Biases in Thinking, Judgment and Memory* (pp. 201-213). New York: Psychology Press.
- Hart, C. (2011). Legitimising assertions and the logico-rhetorical module: Evidence and epistemic vigilance in media discourse on immigration. *Discourse Studies*, 13 (6), 751-769.
- \_\_\_ (forthcoming). Argumentation meets adapted cognition: Manipulation in media discourse on immigration. *Journal of Pragmatics*.
- Henrich, J. and R. Boyd (1998). The evolution of conformist transmission and the emergence of between-group differences. *Evolution and Human Behaviour*, 19, 215-241.
- Hurford, J.R. (2007). *The Origins of Meaning: Language in the Light of Evolution*. Oxford: Oxford University Press.
- Jackson, S. (1996). Fallacies and heuristics. In J. van Benthem, F.H. van Eemeren, R. Grootendorst & F. Veltman (Eds.), *Logic and Argumentation* (pp. 101-114). Amsterdam: Royal Netherlands Academy of Arts and Sciences.
- Maillat, D. (forthcoming). Constraining context selection: On the pragmatic inevitability of manipulation. *Journal of Pragmatics*.
- Maillat, D. and S. Oswald (2009). Defining manipulative discourse: The pragmatics of cognitive illusions. *International Review of Pragmatics*, 1, 348-370.
- Maillat, D. and S. Oswald (2011). Constraining context: A pragmatic account of cognitive manipulation. In C. Hart (Ed.), *Critical Discourse Studies in Context and Cognition* (pp. 65-80). Amsterdam: John Benjamins.
- Mercier, H. & D. Sperber (2009). Intuitive and reflective inferences. In J. St. B. T. Evans and K. Frankish (Eds.), *In Two Minds: Dual Processes and Beyond* (pp. 149-170). Oxford: Oxford University Press.
- Mercier, H. & Dan Sperber (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioural and Brain Sciences*, 34, 57-111.
- Milgram, S. (1974). *Obedience to Authority: An Experimental View*. New York: Harpercollins.

- Origg, G. & D. Sperber (2000). Evolution, communication and the proper function of language. In P. Carruthers & A. Chamberlain (Eds.), *Evolution and the Human Mind: Language, Modularity and Social Cognition* (pp. 140–169). Cambridge: Cambridge University Press.
- Oswald, S. (2011). From interpretation to consent: Arguments, beliefs and meaning. *Discourse Studies*, 13 (6), 806-814.
- \_\_\_ (forthcoming). Rhetoric and cognition: Pragmatic constraints on argument processing.
- Pohl, R.F. (2004). *Cognitive Illusions: A Handbook on Fallacies and Biases in Thinking, Judgement and Memory*. Hove: Psychology Press.
- Sperber, D. (2000). Metarepresentations in an evolutionary perspective. In D. Sperber (Ed.), *Metarepresentation: A Multidisciplinary Perspective* (pp. 117-38). New York: Oxford University Press.
- \_\_\_ (2001). An evolutionary perspective on testimony and argumentation. *Philosophical Topics*, 29, 401-13.
- Sperber, D. and D. Wilson (1995). *Relevance: Communication and Cognition*, 2nd edn. Cambridge, MA: Blackwell Publishers.
- Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origg, G. and D. Wilson (2010). Epistemic vigilance. *Mind and Language*, 25 (4), 359-393.
- Tversky, A. and D. Kahneman (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, 185 (4157), 1124-1131.
- Walton, D. (2006). *Fundamentals of Critical Argumentation*. New York: Cambridge University Press.
- Wilson, D. and D. Sperber (2012). *Meaning and Relevance*. Cambridge: Cambridge University Press.