Defining Manipulative Discourse: The Pragmatics of Cognitive Illusions

Didier Maillat and Steve Oswald

University of Fribourg and University of Neuchâtel, Switzerland
didier.maillat@unifr.ch and steve.oswald@unine.ch

Abstract

Manipulative discourse has attracted a lot of attention in various adjacent domains of linguistic research, notably in rhetoric, argumentation theory, philosophy of language, discourse analysis, pragmatics, among others. We start with a review of the existing definitions provided in these fields and highlight some of the difficulties they encounter. In particular, we argue that there is still a need for an analytic model that makes predictions about manipulative discourse. We propose an alternative account of manipulation couched in the relevance-theoretic framework which treats manipulation as a two-step communicative attempt at misleading the context-selection process when interpreting a target utterance. We argue further that such attempts systematically exploit the inherent weaknesses or flaws of the human cognitive system that are amply discussed in cognitive psychology under the heading of "cognitive illusions". We claim that such a model correctly captures classical instances of manipulative discourse which fall outside the scope of other accounts.

Keywords

argumentation, cognitive illusion, fallacy, manipulation, context selection process, Relevance Theory

Introduction

Although the distinction between verbal manipulation and alternative non-manipulative means by which speakers may get their addressees to engage in a particular course of action or to believe the truth of certain propositions is fairly intuitive, a consistent theoretical model of the former is not available yet. Scholars from disciplines in the Humanities as varied as social psychology, philosophy, anthropology, communication science, argumentation theory, linguistics, psychology and critical discourse analysis have tried to tackle the phenomenon or to explore some of its facets, but consensus, let alone a unified account, has not been reached yet. The reason for this, in our opinion, relates to the heterogeneous nature of manipulation; most of these approaches

indeed focus on some of its particular aspects, but do not – and sometimes cannot – take a step further to consider the phenomenon comprehensively. Answering the "what" question ("what is manipulative communication?") is a very difficult task, in so far as the identification of manipulation as a communicative phenomenon proves to be problematic with regard to a goal of descriptive adequacy. Contemporary approaches, in this line of work, have tried to bring forth several criteria in order to assess what it is that makes a discourse manipulative; yet the question is far from being settled, as we will see further along.

Our proposal aims at reviewing these criteria and explores a relatively new direction for research on manipulation, i.e. a cognitive one. In section 1 we will discuss some of the criteria that have been proposed to characterise manipulation. These are truth-conditionality violation, speaker interest, covertness, social inequality and intention. We will show how none of these may constitute a necessary condition for manipulation, as they might constitute an obstacle to providing an operative criterion in order for a theoretical model to be adequate from a descriptive point of view. Some of these criteria, we will also argue, fail to make the right predictions as they either over-generate or under-generate. Consequently we will suggest that studying this phenomenon may benefit from a shift of perspective. Building on the assumption that manipulation gains its effectiveness through the manipulated individual's processing of the manipulative input, section 2 will review literature on cognitive processing of linguistic information, in order to establish the inherent risk associated to meaning derivation and to defend the idea that manipulation may actually exploit the fact that "errors" are likely to occur in speech processing. In this respect, we take on board Rigotti's idea according to which "the dynamics of manipulation are very close to the dynamics of human error" (2005: 69). This hypothesis will draw on contemporary research in psychology and cognitive pragmatics (e.g. Pohl, 2004; Sperber & Wilson, 1995). Section 3 will in turn develop the idea and sketch out a pragmatic model which accounts for the way manipulators may actually induce such "undue" processing. This alternative analytical approach will allow us to reconsider the nature of manipulative discourse and to provide a new definitional framework to the processes involved in manipulation. In order to do so we propose to integrate the complexity inherently associated with the heterogeneous nature of manipulation by focusing our attention on the processes involved in the actual success of the manipulative attempt (i.e. the cognitive mechanisms involved in speech processing). In other words, we will try to understand the "whatness" of manipulation by answering the "how does it work" question from a cognitive point of view. In doing so we will address

a point recently raised by Cummings (2004: 178ff) who emphasises that while there exists a clear potential for a pragmatic analysis of argumentation phenomena and their governing principles, (neo-)Gricean pragmatic models have not been applied to this domain, which renders "[p]ragmatic studies of argumentative principles […] all the more urgent […]" (Cummings, 2004: 185).

1. The Heterogeneous Nature of Manipulation: Available Clues

Contemporary research on manipulation has usually been concerned with the nature of the manipulative message, its properties, and the social conditions in which manipulation is likely to occur. This section discusses the limits encountered by such accounts which try to isolate the necessary features that make up manipulative discourse. Ultimately, we will propose that the identification of manipulation may not be as informative, from a scientific point of view, as an account of the mechanisms by which manipulation may be successful.

1.1. Truth-Conditions and Felicity Conditions

Truth and felicity conditions have been proposed as relevant criteria defining manipulative strategies in terms of falsity and insincerity (Rigotti, 2005); the main assumption behind this view roughly amounts to considering that some of the manipulative message's features, in terms of the content it encodes, do not comply with the communicative standards of verbal interaction.[1] This is the point made by Rigotti, who suggests that in manipulation, "what is negative has to be somehow disguised as something positive", and that manipulation "twists the vision of the world […] in the mind of the addressee" (2005: 68). Typically, lies would fall under this category, since it is crucial for a lie's success to be taken as a truthful statement. If a child lies to his parents by telling them that he did not break a valuable vase,[2] it might be argued that he has tried to be manipulative. Similarly, insincere promises, such as those politicians might utter in

---

[1] In terms of speech act theory, this amounts to saying that a manipulative utterance generally violates some felicity conditions of the speech act at stake, or, from a Gricean perspective, that it *covertly* violates the Cooperative Principle and possibly the conversational maxims.

[2] The lie itself is merely about providing a false statement while asserting it as a truth; moreover, if the child provides an alternative explanation, by for instance convincingly blaming the dog instead and thus avoiding potential punishment, we could say the child's contribution was manipulative, in so far as it is further constraining the parents' processing of information by providing a plausible context to explain the situation.

support of their future election, can also turn out to be manipulative, as they may break the sincerity conditions which define a promise.[3]

However, it seems difficult to regard a false statement as a necessary or even as a sufficient condition for manipulation to take place, as Rigotti himself (2005) acknowledges. First of all, there is strong evidence in favour of alternative manipulative mechanisms which do not involve lying. Such is the case, for instance, when a speaker provides a plausible justification for a proposition which complies with an actual state of affairs, while the real motivation for her statement is not the one she makes manifest, but one which she wants to keep concealed, while at the same time obtaining the desired effect on the addressee. In this first type, the manipulative strategy relies on shifting the addressee's attention as it were, stating A (which is true) so as to hide the truth of a proposition B which would weaken or contradict the view entertained by the speaker.

Along this line of thought, Parret (1978) argues that lying is limited to cases where a speaker communicates a false proposition which misleads the hearer into entertaining a new belief, while manipulation may function in several different ways: for example, the communication of a proposition may mislead the addressee into giving up a legitimate belief, into reinforcing a problematic belief or into preventing him from entertaining a legitimate belief, all of which can be achieved without resorting to lying on the speaker's part, which would, therefore, be difficult to capture either by the restrictions imposed by the very notion of lying or through an account based solely on the notion of truth. What this shows, beyond Parret's analysis, is that truth turns out to be an insufficient criterion when it comes to characterising manipulative communication.

Secondly, some researchers point out that the manipulative status of a lie may even be questioned in a few cases: the case of parents telling their children about Santa Claus raises this particular issue, as highlighted by Saussure & Schulz when they argue that "it would be counter-intuitive that the parent telling the child about Santa Claus is performing manipulation" (2005: 2). If we are simply looking at the nature of the message parents make manifest to their child by talking about Santa Claus, truth-conditions are indeed violated, and, specifically, the existential presupposition. However, existential presupposition is also violated when parents tell tales about unicorns, and more generally, when a speaker's utterance is about fiction. Although it could be said that there is a difference between Santa Claus and unicorns in that children are

---

[3] In this example, the felicity (and by "unorthodox" extension, its truth) of the statement is a function of the sincerity with which it is uttered, hence the truth-functional violation.

aware that unicorns are to be found in fairy tales only, while they can reasonably be taken to believe in the actual existence of Santa Claus in the real world, to the extent that it might even have a tangible effect on their behaviour (being good all year in view of getting presents at Christmas), there is *a priori* no reason to assume that truth-conditions are responsible for that difference. Be that as it may, it could still be argued that the deliberate use of a false statement is manipulative, as long as the addressee is not aware of it; that is, if we consider that the hearer's processing of information turns out to be biased by virtue of the constraints the speaker imposes on the message. Of course, from a somewhat more ethical and less technical point of view, we might express reservations to call the Santa Claus lie manipulation, as Saussure & Schulz (2005) do, simply because the example seems radically different from one where a cult guru manages to manipulate his followers and gets them to willingly give him large amounts of money, or, more dramatically, to give up their own lives, as was the case a few years ago with the sadly famous Order of the Solar Temple collective suicide. But this is a separate question, and defining manipulation in terms of moral or ethical acceptability would call for discussions that exceed the scope of the present paper, although it does suggest that a discussion about the manipulative status of lies bears technical as well as ethical implications, since it involves far more than unsatisfied truth-conditions.

From a speech-act-theoretical perspective, it could be envisaged to consider that manipulation violates felicity conditions, and among them, the sincerity condition in particular; we mentioned earlier in this respect the scenario of a politician uttering a commissive without actually being committed to the ulterior satisfaction of the propositional content conveyed. However, if one were to define manipulation within a speech-act-theoretic framework, a major problem, discussed by Parret (1978), would emerge: it is the idea that the speech act of manipulation cannot comply with the principle of expressibility (cf. Searle, 1969) because it cannot be translated with an explicit performative such as "I manipulate you + propositional content".[4] Therefore, trying to couch a definition of manipulation in terms of felicity conditions would not take us very far; as Parret (1978) noted: the notion of manipulation as a speech act is either self-destructive or it destroys the conceptual frame itself, were it to be elaborated in terms of speech act theory. That is to say that manipulation is

---

[4] In fact, we believe that making the manipulative intention explicit would make the manipulative attempt collapse altogether. We take this to be in the very nature of manipulation: a manipulative intention is not meant to be communicated nor recognised (see sections 1.3 and 1.5 below).

a kind of speech act which is not meant to be recognised at all, not even indirectly. What these observations indicate is that defining manipulation according to truth or felicity conditions leaves too many holes in the model.

1.2. Speaker Interest

The notion of speaker interest has been highlighted by many researchers as one key feature of manipulation (e.g. van Dijk, 2006; Rigotti, 2005; Saussure & Schulz, 2005). The claim is that manipulation is an intrinsically goal-oriented phenomenon designed to satisfy speaker interest. As van Dijk notes (2006: 360), "manipulators make others believe or do things that are in the interest of the manipulator and against the best interests of the manipulated". Saussure also highlights the connexion between speaker interest and the manipulative attempt:

> […] communication is manipulative when the speaker retains some relevant information, or provides the correct information in order for the hearer to conclude that he should behave in a way which favours the speaker's interests, without being aware of it. (2005: 119-120)

The notion of interest is intimately linked to that of goal. An interest can be defined as a goal that an individual might reasonably be taken to be willing to attain. It should be noted that in communication there are properly linguistic goals (i.e. somewhat "local" goals pursued in the processing of information itself) and extra-linguistic goals ("global" goals, such as the pursuit of happiness, well-being, receiving favours, etc.).[5] Communication, including manipulative communication, cannot take place unless the message conveyed is understood by the addressee. In this sense, it could be regarded as one of the speaker's basic interests to intend that her message be adequately interpreted. Consequently, it would make little sense to use this acceptance of *interest* to describe manipulation. We therefore take it that when researchers discuss the issue of interest satisfaction in manipulation, they do so mostly with extra-linguistic interests in mind.

Now, several problems arise when examining the issue in further detail. First of all, would it be possible for manipulation to favour the addressee's interest? In other words, is it possible to manipulate people to their own advantage? Although this question might receive a negative answer at first sight, we argue that a positive answer should be preferred in a number of cases. Let us take an

---

[5] See Lumsden (2008) for a state-of-the art discussion of the distinction between linguistic and extra-linguistic goals from a neo-Gricean perspective.

example: two friends, one of which is suffering from a severe, potentially terminal illness, are having a conversation about medical treatments. Assume also that the latter has a phobia about needles and cannot stand the thought of being stung. Further assume that his friend just heard on the news that a new injection-based treatment which, in some cases, can cure the disease, and which considerably alleviates pain in all cases, has just been made available at the local hospital. What would we make of the communicative conditions, if the healthy speaker manages to get his sick friend to go to the hospital without disclosing the nature of the treatment? It can hardly be claimed that this is an instance of hearer interest violation, to the extent that both participants in the exchange can reasonably be taken to pursue the satisfaction of the sick hearer's interest (i.e. to get better). Yet, we claim that this is an example of manipulation, notwithstanding the absence of hearer interest violation. And in this case, speaker and hearer interest do reasonably converge, at least in regard to the goal of improving the hearer's health. Of course, one could argue that in spite of this particular convergence, an asymmetrical configuration still remains here, in that the manipulative attempt runs contrary to the hearer's phobia of needles. Yet it seems to us that there is room to consider that it was intended to fulfil a "higher-order" interest; in other words, the success of this particular manipulative attempt involves the subordination of the violation of the secondary "avoid-phobia" interest to the satisfaction of the "improve-health" interest. Here the manipulative goal is foremost to improve the person's health, not to trick him into getting stung in spite of his phobia, and so we would say that in this particular case, the real motivation of the manipulative attempt translates into a convergence of speaker and hearer interest.

People do have conflicting interests, but this does not mean that manipulation is only about violating a hearer's interest: some violations, which in this sense can appear to be incidental, may happen in the pursuit of a goal which, in the end, satisfies one of the hearer's main interests. What we wanted to stress with this example is that the consideration of *both* speaker and hearer antagonistic interests is not necessary to describe manipulation; examples such as that given above illustrate that manipulation, at least in principle, does not necessarily require the satisfaction of speaker interest *at the expense* of the hearer's interest. This counter-example casts doubt on van Dijk's take, according to which manipulation favours speaker interest *and* runs contrary to hearer interest.[6]

---

[6] We could indeed always consider that manipulation runs contrary to at least some of the hearer's interests; however, contrary to a widespread take on manipulation, we do think that it is possible to manipulate people to their own advantage. This suggests that what makes a statement manipulative is not whether it is detrimental in any way to the hearer; we would rather

A second problem with the operability of the notion of interest lies in its relativity. What may be considered as a legitimate satisfaction of interest for someone may not be for someone else. What is more, there seems to be room to question the legitimacy of the researcher's own judgment: on which grounds can we say that a researcher's judgment is accurate when it comes to assessing people's interests? These questions bring up the classical issue of the analyst's own (ideological) bias, and could definitely turn out to be problematic should the notion of interest be a central part of a systematic account of manipulation.

What these two issues point to is that we cannot expect the notion of speaker interest to function on a descriptive level in order to account for manipulation. While we do acknowledge that a component of the definition of manipulation relates to interest satisfaction, we also consider that if the analyst chooses to rely on it on a descriptive level, s/he must be prepared to accept that her/his account might prove difficult to apply analytically, since (i) it may involve ideologically-loaded reasoning, and (ii) it might also face the problem of finding manipulative strategies in every utterance.

1.3. Covertness

A very intuitive and widespread view on manipulation holds that in order to succeed, the manipulative attempt must remain covert. Quite straightforwardly, we could indeed assume that it is likely that manipulation will fail if spotted. Furthermore, there are reasons to doubt that manipulators would explicitly endorse their manipulative behaviour in communication. These preliminary remarks tend to confirm that it is necessary for manipulation to work to go unnoticed.[7]

However, if we are to assess covertness in manipulation, we should first be clear about its scope. A quick review of some elements that may remain covert in a manipulative message shows that many different things can be deemed to be covert:

---

cautiously assume that what makes a statement manipulative is the specific type of processing it calls for (see below). Moreover, if we were to loosen van Dijk's criterion and assume that manipulation violates *some* hearer interest (whatever it may be), the feature would lose its predictive power yielding a model which would overgenerate (since most utterances are bound to run against at least one of the hearer's many interests).

[7] This intuition is actually reflected in natural language. Take the following pair of examples:

  (a) Let me persuade you to come to the cinema with me.
  (b) ?? Let me manipulate you to come to the cinema with me.

While it sounds perfectly natural to utter something like (a), the acceptability of an utterance such as (b), which goes on record, turns out to be problematic, because covertness indeed seems to be

- *the overall manipulative intention*: as noted above, the transparency of the manipulative attempt would seemingly compromise its success, making it crucial for the manipulative intention to remain concealed;
- *local and basic linguistic strategies*: a linguistic device relying for instance on particular pronominal usages (e.g. *us vs. them*) could aim at surreptitiously establishing an in-group/out-group ideological discrimination, which in turn may influence further beliefs and behaviour;
- *global strategies*: these strategies aim at creating adequate psychological and social conditions, such as controlled peer pressure or strategies designed to increase the faith in the speaker;
- *discursive strategies*: a complex articulation of speech acts fulfilling a specific function, e.g. a fallacious argumentative move (see van Eemeren & Grootendorst, 2004), may also require being covert in order to succeed.

We see that a great many aspects of verbal communication can remain covert. This complexity may compromise the operability of the criterion of covertness in an account of manipulation, insofar as different instances of manipulation may display different types of covertness. Such an enterprise runs the risk of being difficult to circumscribe in a model, given the heterogeneity of the set of variables. Some refer to psychological and cognitive phenomena (speaker intentionality), others to social issues (peer pressure, group behaviour), and others are about particular linguistic or discursive mechanisms (local and global strategies). In other words, the major difficulty we face in the applicability of covertness as a distinctive feature is a consequence of its problematic definitional range.

One way out of the problem would be to consider that these multifarious aspects of covertness in communication can be addressed through an examination of the particular processing the manipulative message calls for, since, in the end, their effect – or power – is only realised when an addressee processes the information conveyed by the manipulative argument. Also, a list like the above points to the need for multidisciplinary efforts in the study of argumentative phenomena like manipulation as it ranges over psychological, sociological, linguistic, pragmatic, and even anthropological parameters. Such an enterprise is starting to develop; Clément (2006), for instance, reconciles sociological, anthropological and cognitive aspects of belief fixation, and credulity in particular; however, to our knowledge there has been no attempt to

---

crucial in manipulation (unless, of course, we take (b) to be meta-linguistically intended to produce a humorous effect). In Gricean terms (Grice, 1989), this observation amounts to saying that manipulation is an instance of Gricean "*unostentatious violation*", (1989: 30).

integrate these various aspects into one consistent model. It is our purpose to sketch out the lines along which such an enterprise could be carried out within a pragmatic framework.

But for the time being, and coming back to the topic of covertness in manipulation, let us mention a further problem we would face, and which emerges from psychological evidence (e.g. Barton & Sanford, 1993; see in section 3 below the discussion of the Moses illusion). The results obtained through experimental designs to test the extent to which specific, crucial, lexical information is or is not accessed by subjects while interpreting a fairly simple narrative text revealed that even when they are warned about a potentially manipulative content (i.e. when they are told for instance that there are anomalies in the text), the anomalies may still go unnoticed, and the manipulative effect would still be achieved. This kind of observation would support the idea that covertness may not even be a necessary condition of manipulation, since in cases where you neutralise it by making addressees aware of the manipulative nature of the utterance, manipulation can still take place. While it could be said that the text did not explicitly alert the subjects to the lexical nature of the anomaly, we see nevertheless that a higher degree of awareness to covert strategies is not sufficient to undermine manipulative attempts. It remains to be seen how a more precise warning would have led to a failure of the manipulative intent.

All in all, manipulation indeed seems to rely on covertness most of the time, but the notion remains difficult to capture within a theoretical model and proves somewhat shaky to be systematically integrated in a descriptive account, at least if we wish to use it as a discriminatory criterion, as many other types of speech acts also require covertness (e.g. lies). While we wish to maintain that manipulation must be covert, the very difficulty involved in systematically establishing the nature, role and scope of covertness in manipulative communication makes it a difficult feature to integrate in a descriptive model.

1.4. Social Inequality

Research in Critical Discourse Analysis (e.g. van Dijk, 2006) and in psychosocial science (e.g. Milgram, 2004; Cialdini, 2009) stresses the importance of the social conditions manipulative communication builds on. According to van Dijk, manipulators "need to satisfy personal and social criteria that enable them to influence others in the first place" (2006: 362), whereby it is assumed that asymmetrical social conditions of interaction are required for manipulation to take place. Typically, this is illustrated by an asymmetry between speaker

and addressee(s) in hierarchical social positions (parent/child, professor/pupil, politician/member of the public, etc.), knowledge availability (highly educated/poorly educated people, scientist/non-scientist, expert/non-expert, etc.), access to public discourse, and so on. Underlying this view is the idea that manipulation exploits these forms of social asymmetry and the associated imbalance in the relations of power and domination in favour of the manipulator.

While we do not deny that social parameters indeed contribute to successful manipulation, there are reasons to doubt that these are necessary conditions. They certainly make explicit certain conditions under which manipulation occurs, but in our view they are not powerful enough to identify all cases of manipulation. The justification for this is that we see no compelling reasons to straightforwardly exclude that manipulation could occur in the absence of prior social domination: in principle, we do not see why children could not manipulate their parents, why friends could not manipulate each other, why pupils could not manipulate their professors, etc. More importantly, on the one hand, a socially powerful individual or group may very well decide not to use manipulation when attempting to convince another, subordinate group; on the other hand, manipulation can obtain within reverse social power relations. In other words, social asymmetry does not *necessarily* entail manipulation. Van Dijk is aware of the difficulty but dismisses these cases on the grounds that they hinge upon personal psychological factors:

> I limit my analysis to social criteria, and ignore the influence of psychological factors, such as character traits, intelligence, learning, etc. In other words, I am not interested here in what might be a 'manipulating personality', or in the specific personal way by which people manipulate others. (van Dijk, 2006: 362)

It is interesting to highlight the fact that van Dijk both recognises and dismisses a non-social component in manipulative strategies since he identifies two aspects of manipulation, that is, its social and psychological underpinnings. We propose that a suitable model for manipulative discourse should try to cater for both aspects.

The same quote also reveals that his conception of manipulation is *speaker-oriented*, for he associates the psychological feature of manipulation to the personality of the speaker. We contend that while it is desirable to include the psychological component of manipulation, the latter is not reducible to the speaker's personality. We argue further that the crucial psychological aspect of manipulation lies in the way it constrains the addressee's interpretative processes. In this respect, we propose that an appropriate pragmatic model of manipulation must shift its focus on and account for the cognitive (psychological) processes which underlie the interpretation of manipulative discourse.

That is to say that we propose a definition of manipulation which is *addressee-oriented*. However, this does not rule out the relevance of social factors in the suggested account. Rather, we want to consider how these – together with other parameters – may effectively constrain the addressee's processing of the manipulative utterance.

In doing so we wish to avoid van Dijk's more restricted view of manipulation – expressed below – in order to be able to account for a wider range of cases:

> […] if manipulation is a form of domination or power abuse, […] it only makes sense to speak of manipulation, as defined, when speakers or writers are manipulating others in their role as a member of a dominant collectivity. (van Dijk 2006: 364)

We will see in section 2, how a more comprehensive view of manipulative communication can extend van Dijk's account in order to be able to deal with cases of manipulation that go beyond relations of social dominance.

1.5. Intention

Another speaker-oriented parameter which has been regarded as relevant to research on manipulation, though a highly problematic one from a descriptive point of view, is intentionality. Is manipulation intentional? Providing a negative answer to this question amounts to considering that people may accidentally manipulate others. Such a position may turn out to be rather hard to defend: in addition to having to deal with the difficulty involved in accounting for an accidentally-occurring phenomenon, such a line would also leave us in a theoretical dead end as it would rule out the notion of covertness discussed above and assumed by the majority of researchers in the field, as we would have to explain how a speaker can hide something she is not even aware of. We will therefore assume that there is some deliberate deceptive intention involved in manipulation.

The centrality of a deceptive intention in manipulation bridges the gap between theories of argumentation and pragmatic accounts which consider intention as a central aspect of communication (particularly in the field of post- and neo-Gricean pragmatics, e.g. Carston, 2002; Levinson, 2000; Sperber & Wilson, 1995). The latter may thus represent an interesting option for an account of manipulation (as argued more generally for argumentation by Cummings (2004: 178ff)). Specifically, manipulation proceeds like any other communicative exchange as an attempt on the part of the addressee to retrieve the speaker's intention, but relies crucially on a mismatch between speaker intention and the intention attributed to her by the addressee. Since manipulation is taken to be covert, it would be paradoxical to assume that it is based on an implicature derived from an intention which is, by definition,

not meant to be communicated, let alone recognised; this would be inconsistent with the core assumptions underlying contemporary pragmatic research.

Given that the persons who are the targets of manipulation are not aware that they are being manipulated, we assume that their processing of the speaker's utterance is completely straightforward, and in particular that hearers assume speaker cooperativeness (in the Gricean sense). In other words, from the addressee's point of view, the type of processing involved in cases of standard, cooperative, communication and in cases of manipulation is the same: the difference is rather to be found at the speaker's end of the communicative process. This is also the conclusion Attardo (1997) arrives at. He states that in competitive modes of interaction (among which we find manipulation), "parties exploit the assumption of cooperation to further their unilateral goals" (1997: 778).

Furthermore, we claim that the focus of the account should be on the cognitive processes of interpretation. The issue such an approach will have to deal with is therefore not so much an assessment of the non-recovery of the manipulative intention (inasmuch as the latter does not have a *communicative* status) than an assessment of how it is that the hearer's mental processing of information, in manipulative communication, takes a somewhat "undue" path. As pointed out by Rigotti (2005: 69), "the dynamics of manipulation are very close to the dynamics of human error". It is precisely this assumption that will underlie our discussion of manipulative communication. Our claim is that manipulation exploits the way our mechanisms of information processing work; that is, a necessarily imperfect and biased way.

Yet, there are at least two counts on which an addressee-oriented model which does not depend, at a descriptive level, on the notion of a deceptive intention can provide a serious alternative when accounting for manipulation:

i) how could intention-based approaches capture cases of second-hand manipulation, where a credulous – already successfully manipulated – disciple repeats a manipulative creed? In other words, how can they handle cases where the result of the exchange may turn out to be the same as the result of manipulation, when there was no actual manipulative intention on behalf of the speaker?

ii) how would they account for cases of manipulation where the manipulator's goal is merely to distract the hearer from relevant information? These cases, an example of which we will see further on (see section 4 below), do not have an actual manipulative content, because the speaker is not aiming at conveying a particular set of representations, but rather at preventing the hearer from forming specific representations.

The example of second-hand manipulation might be regarded as controversial in so far as its manipulative nature is concerned. Can we still consider it as

a proper instance of manipulation? Two answers can be provided to this thorny issue: first, it could be considered that the manipulated speaker, who propagates the manipulative speech, is relaying the original information, and thus it is as if the original ("first-hand") manipulative intention is indirectly relayed through the disciple's speech though not intended as such by her. Second, if we define manipulation as a set of conditions limiting the interpretative processes of the addressee instead of a property of the communicative intention, we find ourselves in a position to capture the intuitively correct fact that whoever listens to the disciple can indeed be *manipulated*.

We cannot think of any strong reasons to refute the hypothesis that the same (originally manipulative) message will possibly yield the same representations, even if produced by a speaker with no manipulative intention. Since a given hearer will process a manipulative message following the standard cooperative procedure, the idea is not to consider the manipulative message from the perspective of its potential distinctive features (such as the manipulative intention), but rather from the perspective of the particular type of processing it calls for. In this sense our definition tries to capture the cognitive experience associated with manipulation.

In order to explore this new direction, as we will see in the following section, we will postulate that manipulation can be best defined in terms of the constraints it imposes on mental processing: the hypothesis we will defend consists in shifting the focus from the type of concern evoked throughout the previous subsections to conditions affecting the reception of a manipulative message; specifically, we will define manipulation as a set of constraints limiting the processes of contextual selection. Instead of concentrating on the features of the manipulative message by trying to characterise it from the perspective of the manipulator, we will reverse the angle and consider manipulation from the perspective of the manipulated. We will thus investigate how manipulated hearers are (mis)led to process the information contained in the message in a somewhat restricted way. This move, we claim, will do away with the complexity expounded above in the discussion of the defining criteria of manipulation, while allowing us to determine a set of recurrent features which characterises manipulation in a large variety of cases.

2. An Alternative Account

We argue that a more appropriate and explanatorily more powerful account of manipulation can be couched in the framework of Relevance Theory. In the remaining sections we will show how a new theoretical model of manipulation can be captured in relevance-theoretic terms, and how the latter can help us

understand both the more standard forms of manipulative discourse, as well as the more marginal forms.

According to Sperber & Wilson (1995), human communication relies on the assumption that all utterances are optimally relevant within the specific context in which they are produced. Optimal relevance is understood as a cognitive state which corresponds to an optimal ratio between the cognitive efforts required to process the utterance, and the cognitive effects yielded by such an interpretative process. Our account of manipulative discourse proposes that manipulation takes advantage of and exploits the cognitive dynamics which underlie this mechanism. Crucially, whereas many accounts assume that manipulation is a form of inherently flawed communicative activity, we argue that, from an interpretative perspective, manipulative discourse relies on the same context-construction or context selection procedure as in the case of a "normal" utterance. The communicative twist introduced by manipulative discourse lies in the external restrictions or constraints imposed by the manipulator on the interpretative process, even though the process itself remains unchanged.

This argument is to be related to another claim made by Allott & Rubio Fernández (2002) who show that the interpretation of certain lexical items can lead to "cognitive shortcuts" whereby the addressee only activates the most salient conceptual assumptions related with a lexical entry, a phenomenon known as *shallow processing*. Shallow processing gives rise to so-called *ad hoc concepts*, that is to say concepts which do not correspond to the fully-fledged set of conceptual assumptions that correspond to a full exploitation of the concept but to a contextually tailor-cut, targeted interpretation.

According to Allott & Rubio Fernández (2002), this cognitive tendency displayed during interpretation sheds a new light on phenomena such as the Moses illusion[8] by providing a systematic account to cases where cognitive processing can lead to an erroneous/misled interpretation.

In order to address the initial question "what is manipulation?", we want to posit that manipulation corresponds to a communicative strategy that relies on these exact same processes and exploits the cognitive strategies deployed during interpretation in order to optimise the use of resources. In other words, manipulative discourse is a form of communication that puts the addressee in a situation where s/he will be led to shallow-process contextual assumptions. Manipulation, therefore, constitutes a form of cognitive constraint on the selection of contextual assumptions.

---

[8] See section 3 below.

In the proposed model, manipulative discourse functions as a twofold process that first puts a strong constraint on the selection of contextual assumptions which are accessed to interpret a target utterance *U*. This first constraining element ensures that the target utterance is interpreted within a limited context, *C*, and – most importantly – it ensures that any alternative set of contextual assumptions, *C′*, is *not* accessed. That is to say that manipulative discourse is a form of communicative attempt at blocking the context selection process described by Sperber & Wilson (1995) in the following quotation, and which is at the heart of interpretation:

> Humans […] try to process information as productively as possible; that is, they try to obtain from each new item of information as great a contextual effect as possible for as small a processing effort. The assessment of relevance is not the goal of the comprehension process, but only a means to an end […].
>
> If this is true, it suggests a complete reversal of the order of events in comprehension. It is not that first context is determined, and then relevance is assessed. […] it is relevance which is treated as given, and context which is treated as a variable. (141-142)

Coming back to what we proposed before, manipulation is reanalysed as an attempt to limit and, hence, misguide the latter process of context selection in order to control the comprehension of some target utterance. While we recognise two different components in a manipulative use of language, they do not need to be distinct discursively and can be combined within a single utterance as we will see later on when we look more closely at an actual example. Nevertheless, it is important to distinguish between on the one hand, the constraining and inherently manipulative move and the (often) innocuous target utterance whose interpretation is being constrained on the other hand.

Interestingly, such an account is capable of explaining both standard cases of manipulative discourse as well as cases of "second-hand manipulation" (see above). As an example of the latter type of manipulation, we can think of a situation where the creed of a revisionist group is being relayed by some of its disciples even though they are not aware of the controlling – and therefore misleading – attempt which is built into it. In other words, a speaker can manipulate others unaware, as the disciples mentioned here would. To put it differently, as manipulation is now defined as a blocking action on context selection procedures during the interpretation of an utterance, it follows that the effect can be triggered unintentionally.

As was already noted before, while we couch our theoretical proposal within a pragmatic model, we take manipulation to be essentially a cognitive process, which successfully misleads context selection. As a result, in line with the views expressed in the RT literature, we make the claim that manipulation transcends

language and is best explained at a more fundamental level, namely cognition. In this respect, there is a strong adequacy between the cognitive framework used to ground Relevance Theory and the claim we are making here. Furthermore, the cognitive line taken in this paper also allows us to explore some findings made in adjacent domains and open new perspectives on modelling manipulative discourse. We think in particular of the experimental work carried out in the field of cognitive psychology (see Pohl, 2004).

The different papers collected in Pohl (2004) show that human cognition is exposed to cognitive illusions, i.e. situations in which cognitive processes are being misled by a variety of factors. Moreover, it is shown how these "fallacy and bias effects" can be experimentally tested and replicated. In our account, manipulative discourse is assumed to be a linguistic reflection of these underlying cognitive bias effects.

We already touched upon the so-called Moses illusions (see e.g. Allott & Rubio Fernández, 2002) in which an addressee is tricked into answering a question and fails to notice a referential "flaw" in an interrogative utterance. Consider the following utterance:

(1) How many animals of each kind did Moses take on the Ark? (adapted from Reder & Kusbit, 1991)

When asked to tackle such a question most people will answer "two" and miss the anomaly which renders (1) if not meaningless at least trivial. Namely, Moses was never involved in the biblical episode of the Flood. Noah was (Park & Reder, 2004).

In an experimental design testing the impact that (1) and other similar illusions have on people, Erickson & Mattson (1981) found that up to two thirds of the subjects who were tested would miss the anomaly and be tricked into answering "two". As discussed in Allott & Rubio Fernández (2002) and more recently in Park & Reder (2004), several arguments can be put forward that could account for such misled interpretations. Allott & Rubio Fernández (2002) argue strongly for a pragmatic explanation of these phenomena and propose that they are traces of a form of *shallow processing* that ensues logically from the general principles governing human communication and in particular the principle of relevance.

3. Contextual Selection and Shallow Processing: Insights from Relevance Theory

Let us look at the kind of descriptions and predictions achieved by such a model. As we saw, Relevance Theory (see Sperber & Wilson, 1995) provides

a model for standard inferential communication which draws on the assumption that speaker meaning is arrived at through the contextualisation of the utterance that is being interpreted. Contextualisation refers to the construction of the appropriate context. This process is governed by a principle of optimal relevance whereby the addressee tries to maximise the cognitive effects he can generate from the utterance while minimising the cognitive efforts he has to put into the construction of the appropriate context of interpretation. They write that

> [h]umans are not in the business of simply assessing the relevance of new informa-
> tion. They try to process information as productively as possible; that is, they try
> to obtain from each new item of information as great a contextual effect as possible
> for as small a processing effort. The assessment of relevance is not the goal of the
> comprehension process, but only a means to an end, the end being to maximise the
> relevance of any information being processed. (Sperber & Wilson, 1995: 141)

Furthermore, contextualisation is regarded as an incremental process by which new layers of contextual assumptions are accessed if – and only if – the processing of the target utterance has not reached a level of optimal relevance. That is to say that contextualisation, or context selection, will gradually add to the set of contextual assumptions brought to bear in the interpretation of the target utterance, in a stepwise manner, provided a possible point of optimal relevance has not been reached.

Sperber and Wilson (1995) point out that this conceptualisation of utterance interpretation assumes an inclusion relation between different possible sets of contextual assumptions, and crucially they argue that this inclusion relation corresponds to an accessibility relation. In other words, the relevance-theoretic framework implicitly posits that if you can make some contextual assumptions more accessible, they will be more likely to be part of the final contextual selection set than others.

Thus, coming back to the definition of manipulation, in this framework, a manipulative speaker will be taken to increase the accessibility of a certain subset of contextual assumptions, C, within which the target utterance will be almost inevitably processed (from a cognitive perspective), so as to ensure that the addressee does not process the target utterance within a larger context, C', in which it might appear inconsistent or contradictory with some prior background knowledge he upholds. The crucial element in this analysis of manipulative discourse lies in the manipulator's attempt to prevent the addressee from accessing some other less accessible contextual assumptions.

If we go back to the findings put forward by Erickson & Mattson (1981), we realise that the proposed model explains their findings in a straightforward manner. When they analyse the Moses illusions data, Allott & Rubio Fernández (2002) argue that the addressee will only summon some contextual assumptions when

interpreting (1), thereby shallow constructing an *ad hoc concept* in which only some (minimal) assumptions associated with *Moses* are activated, for example:

(2) MOSES*: Biblical figure,

as opposed to

(3) MOSES: Person who led the people of Israel out of Egypt.

While the notion of an *ad hoc concept* is interesting to understand the outcome of the interpretation, our model focuses on the process that leads to such shallow interpretation. In this instance, the cognitive processing of Moses illusions is affected by the fact that the addressee/subject is busy performing a question-answering task and is therefore assessing relevance in connection with the task at hand. Also as the experiment exploits a playful setting in which the subject is challenged in a way, his cognitive effort is concentrated on providing an answer which overrides possible contextual discrepancies. Interestingly, while in most cases the cognitive shortcut described goes unnoticed by the subject, in some instances people exposed to Moses illusions will report spotting the inconsistency and consciously overriding it in order to complete the task. The kind of minimal contextual assumption we are alluding to is echoed in Sperber & Wilson's (1995: 265) revision of their framework, when they distinguish a category of *positive cognitive effect* which is characterised as "contribut[ing] positively to the fulfillment of cognitive functions or goals". In this case, the *ad hoc concept* constitutes such a positive cognitive effect in relation to the fulfillment of the discursive goal, i.e. answering the question.

Coming back to our model of manipulative discourse, the Moses illusions are a good example of a form of utterance that both prevents the addressee from accessing certain contextual assumptions and ensures that the target utterance – (1) itself in this instance – is interpreted within that limited set of contextual assumptions (a set which includes MOSES* but leaves out MOSES). As we can see the suggested account explains in a straightforward fashion well-known manipulative processes by means of a cognitively grounded analysis. It is also equipped to handle recent infamous examples of manipulation that we are all familiar with.

4. A Concrete Example

A striking case of manipulation that took place in the US in the aftermath of the 9/11 attacks against the World Trade Centre and the Pentagon conveniently attracted a lot of attention from the media. It puts in broad daylight

manipulative practices which would typically go unnoticed. The whole affair made the headlines in 2003 and again in 2005 through the Libby scandal.[9] In 2002, Valerie Plame's husband, Joseph Wilson, worked for the CIA and was part of an effort, initiated by Vice-president D. Cheney, to find evidence that Iraq was somehow involved in terrorist threats against the US in connection with the 9/11 attacks. Wilson, a former US ambassador, was one of the persons hired to investigate a lead which linked Iraq with Niger in an attempt to buy uranium for nuclear weapons. Wilson's investigation dismissed the lead as dubious. However, and crucially in our attempt to understand manipulative strategies, the then-president G.W. Bush in his State of the Union address on 28 January 2003 claimed that "the British government has learned that Saddam Hussein recently sought significant quantities of uranium from Africa". That is to say that the intelligence gathered on these allegations had been plainly ignored by the government. This prompted Wilson to make a public comment about the misleading use of available intelligence on these issues. A series of retaliating actions orchestrated by one of D. Cheney's aides (Libby) led to a full exposure of the affair in the press.

In view of what was suggested before, the manipulative technique used here can be analysed in a transparent way. The US government needed a case to go to war against Iraq and they manipulated the American public into interpreting the relevance of a war against Iraq within a very specific limited context: that of terrorist attacks against America. In this instance, the strategy used consisted in constraining the set of contextual assumptions summoned for the interpretation of the target utterance (4):

(4) We must wage war against Iraq

As is obvious from the synopsis provided above, we see that the first manipulative step consists in establishing a link between terrorist attacks and Iraq, thereby limiting the context selection process to $C=\{9/11\}$. In the wake of 9/11, the US government was actively trying to establish terrorist threats as the relevant context concerning Iraq.

Interestingly, by the time the second US-led invasion of Iraq started, a nationwide survey showed that roughly half of the American population thought that Saddam Hussein was connected with the 9/11 events. In other words, the strategy had been extremely successful.

In this prototypical example, the theoretical model proposed for manipulation applies straightforwardly, as we notice an initial attempt to constrain the

---

[9] A good review of the events is available from the *Washington Post* at http://www.washingtonpost .com/wp-dyn/content/graphic/2005/10/29/GR2005102900242.html).

context of interpretation of a target utterance, effectively blocking access to a context C' in which Iraq is *not* involved in any terrorist activity of the kind that led to 9/11. The deliberate attempt to constrain context selection is patent in Bush's statement which ignores the conclusions of Wilson's report.

The cognitive explanation of such a strategy relies on the special status that a set of contextual assumptions such as those associated with 9/11 has from a cognitive point of view. Pickrell et al. (2004: 352-353) explain, using 9/11 as a case in point, that some contexts constitute "highly salient memories, [...] or memories that some theorists have dubbed 'flashbulb memories' for their highly emotional, meaningful, and subjectively permanent nature". By linking Saddam Hussein to terrorist attacks like 9/11, the manipulator tries to ensure that the target utterance U = *We must wage war against Iraq* is interpreted in, *C*, the constrained, emotionally overwhelming flashbulb context of 9/11.

More generally, this example shows that a key to context selection constraint, and therefore manipulation, relies on salience. Typically, the strategy adopted by a manipulator in order to prevent the addressee from accessing an extended set of contextual assumptions, C', makes use of highly salient contextual assumptions which will ensure that a degree of optimal relevance is reached before an extended context is constructed. Flashbulb contexts achieve this particular effect of course, but we do expect the generalisation to hold across other types of highly salient contexts. In the relevance-theoretic framework, we expect two forms of high salience. First, there will be contextual assumptions which require less effort to be accessed. Second, some contextual assumptions will yield greater cognitive effects. In both cases, a subset C consisting of precisely these assumptions will be more salient and therefore function as a context selection constraining device.[10]

Finally, coming back to our earlier example of second-hand manipulation, we see how a credulous disciple can still be indirectly manipulative, in spite of his not having the manipulative intention which is part of the initial strategy. Our account readily allows for an analysis of second-hand manipulation as it does not rely directly on the communicative intention. Thus, a credulous disciple can effectively manipulate other addressees by producing the same contextual constraint as the initial manipulator in spite of the fact that the former might not be aware of the availability of another extended set of contextual assumptions.

---

[10] Maillat & Oswald (forthcoming) explore the various dimensions that the general strategy described here can take.

5. Conclusion: Cognitive Optimism

Sperber, Cara & Girotto (1995) propose an interesting view of the cognitive underpinnings of the quest for optimal relevance. In particular, they convincingly argue that there is an inherent flaw, in a sense, in the cognitive processes which govern the construction of an appropriate set of contextual assumptions. More precisely, in order to increase the performance of the interpretative system, humans tend to take cognitive shortcuts. Sperber, Cara & Girotto write that

> […] people are nearly-incorrigible "cognitive optimists". They take for granted that their spontaneous cognitive processes are highly reliable, and that the output of these processes does not need re-checking. (1995: 90)

Our pragmatic approach of manipulation shows that it is precisely this kind of cognitive optimism which is exploited by manipulative strategies. Specifically, manipulation is an attempt at misleading the addressee which ensures that only a limited set of contextual assumptions is accessed by taking advantage of her/his cognitive optimism.

To conclude, we hope to have shown that such an approach offers a potentially greater descriptive and explanatory power than other traditional accounts in capturing manipulative strategies. Also, this account seems capable of handling cases and manipulative scenarios which do not fit into other models. More importantly, our proposal anchors a pragmatic model of manipulation in the wider field of cognitive psychology and provides an experimentally testable cognitive hypothesis which opens up new perspectives in our understanding of manipulative discourse.

References

Allott, Nicholas and Paula Rubio Fernandez. 2002. This paper fills a much-needed gap. In P. Afuta, A. El Ghali and F. Toussenel (eds.), *Actes de l'Atelier des doctorants en linguistique*, 97-102. Université Paris 7.

Attardo, Salvatore. 1997. Locutionary and Perlocutionary Cooperation: The Perlocutionary Cooperative Principle. *Journal of Pragmatics* 27: 753-779.

Barton, Stephen B. and Anthony Sanford. 1993. A case study of anomaly detection: shallow semantic processing and cohesion establishment. *Memory and Cognition* 21: 477-487.

Carston, Robyn. 2002. *Thoughts and Utterances: The Pragmatics of Explicit Communication*. Oxford: Blackwell.

Cialdini, Robert. 2009. *Influence: Science and Practice* (5th edition). Boson: Pearson / Allyn and Bacon.

Clément, Fabrice. 2006. *Les mécanismes de la crédulité*. Genève: Droz.

Cummings, Louise. 2004. *Pragmatics*. Mahwah, NJ: Lawrence Erlbaum.

Erickson, Thomas A. and Mark Mattson. 1981. From words to meaning: a semantic illusion. *Journal of Verbal Learning and Verbal Behavior* 20: 540-552.

Grice, Herbert P. 1989. *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.

Levinson, Stephen. 2000. *Presumptive Meanings: The Theory of Generalized Conversational Implicature*. Cambridge, MA: MIT Press.

Lumsden, David. 2008. Kinds of conversational cooperation. *Journal of Pragmatics* 40: 1896-1908.

Maillat, Didier and Steve Oswald. forthcoming. Constraining context: a pragmatic account of cognitive manipulation. In C. Hart (ed.), *Critical Discourse Studies in Context and Cognition*. Amsterdam: John Benjamins.

Milgram, Stanley. 2004. *Obedience to Authority: An Experimental View*. New York: HarperCollins.

Park, Heekyeong and Lynne M. Reder. 2004. Moses illusion: Implication for human cognition. In R. Pohl (ed.), *Cognitive Illusions*, 275-291. Hove: Psychology Press.

Parret, Herman. 1978. Eléments d'une analyse philosophique de la manipulation et du mensonge. Prépublication. *Documents de travail et prépublications* 70: 1-33. Centro Internazionale di Semiotica e di Linguistica. Universita di Urbino.

Pickrell, Jacqueline E., Daniel Bernstein and Elizabeth Loftus. 2004. Misinformation effect. In R. Pohl (ed.), 345-361.

Pohl, Rüdiger (ed.). 2004. *Cognitive Illusions: A Handbook on Fallacies and Biases in Thinking, Judgment and Memory*. New York: Psychology Press.

Reder, Lynne M. and Gail Kusbit. 1991. Locus of the Moses illusion: imperfect encoding, retrieval, or match? *Journal of Memory and Language* 29: 633-654.

Rigotti, Edo. 2005. Towards a typology of manipulative processes. In L. de Saussure and P. Schulz (eds.), 61-83.

Saussure, Louis de. 2005. Manipulation and cognitive pragmatics: preliminary hypotheses. In L. de Saussure and P. Schulz (eds.), 113-145.

Saussure, Louis de and Peter Schulz. 2005. Introduction. In L. de Saussure and P. Schulz (eds.), 1-14.

Saussure, Louis de and Peter Schulz (eds.). 2005. *Manipulation and Ideologies in the Twentieth Century: Discourse, Language, Mind*. Amsterdam: John Benjamins.

Searle, John. 1969. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.

Sperber, Dan and Deirdre Wilson. 1995. *Relevance: Communication and Cognition* (2nd ed.). Oxford: Blackwell.

Sperber, Dan, Francesco Cara and Vittorio Girotto. 1995. Relevance Theory explains the Selection Task. *Cognition* 57: 31-95.

Van Eemeren, Frans H. and Rob Grootendorst. 2004. *A Systematic Theory of Argumentation: The Pragma-Dialectical Approach*. Cambridge: Cambridge University Press.

Van Dijk, Teun. 2006. Discourse and manipulation. *Discourse & Society* 17: 359-383.