

Sciences pour  
la communication

# Rhétorique et cognition Rhetoric and Cognition

Perspectives théoriques et stratégies persuasives  
Theoretical Perspectives and Persuasive Strategies

Thierry Herman et Steve Oswald (éds)

Peter Lang

Ce volume met l'accent sur le lien entre démarches cognitives et art du discours qui a toujours été un des enjeux de la rhétorique. Les contributions réunies dans le présent ouvrage n'ajoutent pas une nouvelle couche à l'examen critique des sophismes : en effet elles n'ont pas pour but de dénoncer, mais de décrire le fonctionnement et les effets cognitifs *hic et nunc* de schèmes argumentatifs susceptibles d'apparaître comme fallacieux. Quels sont les mécanismes qui expliquent la «performance» des arguments réputés fallacieux? Comment fonctionnent les stratégies rhétoriques à l'intersection entre cognition, sciences du langage et société?

This volume gathers contributions from two disciplines which have much to gain from one another – rhetoric and cognitive science – as they both have much to say in the broad realm of argumentation studies. This collection neither condemns the fallacious effects of specific argument schemes nor adds yet another layer to fallacy criticism, but studies how argumentation and fallacies work, *hic et nunc*. What are the linguistic and cognitive mechanisms behind the «performance » of fallacious arguments? How do rhetorical strategies work at the interface of cognition, language science and society?

Thierry Herman is a senior lecturer at the Universities of Neuchâtel and Lausanne. Trained in textual linguistics and discourse analysis, his research concerns the use of rhetorical techniques in political, media and academic discourse. Founding member of the Collectif Romand de Recherches en Argumentation (CoRRaA), he is interested in the linguistic form of argumentative schemes and their persuasive effects.

Steve Oswald is a postdoctoral fellow at the Universities of Fribourg and Neuchâtel. His dissertation (2010, University of Neuchâtel) investigated uncooperative and manipulative communication. His current research explores the interface between linguistic pragmatics, argumentation, discourse analysis and cognitive science. He is also a founding member of the CoRRaA and he studies the relationship between fallacious argumentation and its cognitive counterpart for the elaboration of a cognitive account of rhetoric.

Rhétorique et cognition  
Rhetoric and Cognition



## Sciences pour la communication

Vol. 112

### **Comité scientifique**

- D. Apothéloz, Université de Nancy 2  
J.-P. Bronckart, Université de Genève  
P. Chilton, Université de Lancaster  
W. De Mulder, Université d'Anvers  
J.-P. Desclés, Université Paris-Sorbonne  
F.H. van Eemeren, Université d'Amsterdam  
V. Escandell-Vidal, UNED, Madrid  
F. Gadet, Université de Paris Ouest Nanterre La Défense  
J.-M. Marandin, CNRS et Université Paris-Diderot  
F. Martineau, Université d'Ottawa  
M. Milton Campos, Université de Montréal  
J. Rouault, Université Stendhal (Grenoble 3)

Les ouvrages publiés dans cette collection ont été sélectionnés  
par les soins du comité éditorial, après révision par les pairs.

Collection publiée sous la direction de  
Marie-José Béguelin, Alain Berrendonner,  
Didier Maillat et Louis de Saussure

Thierry Herman et Steve Oswald (éds)

# Rhétorique et cognition Rhetoric and Cognition

Perspectives théoriques et stratégies persuasives  
Theoretical Perspectives and Persuasive Strategies



PETER LANG

Bern • Berlin • Bruxelles • Frankfurt am Main • New York • Oxford • Wien

**Information bibliographique publiée par «Die Deutsche Nationalbibliothek»**  
«Die Deutsche Nationalbibliothek» répertorie cette publication dans la «Deutsche National-  
bibliografie»; les données bibliographiques détaillées sont disponibles  
sur Internet sous <http://dnb.d-nb.de>.

Publié avec le soutien du Fonds national suisse de la recherche scientifique  
dans le cadre du projet pilote OAPEN-CH

Ouvrage publié avec le soutien de la Faculté des lettres et sciences humaines  
de l'Université de Neuchâtel

Graphic: Atelier 4b, Sandra Meyer

ISBN 978-3-0343-1547-0 pb.

ISBN 978-3-0352-0271-7 PDF

ISSN 0933-6079 pb.

ISSN 2235-7505 PDF

DOI 10.3726/b10582

ISBN 978-3-0351-9541-5 EPUB

ISBN 978-3-0351-9540-8 MOBI

This book is an open access book and available on [www.oapen.org](http://www.oapen.org) and  
[www.peterlang.com](http://www.peterlang.com).

This work is licensed under the Creative Commons Attribution-NonCommercial-  
NoDerivs 4.0 which means that the text may be used for non-commercial purposes,  
provided credit is given to the autor. For details go to [http://creativecommons.org/  
licenses/by-nc-nd/4.0/](http://creativecommons.org/licenses/by-nc-nd/4.0/)



© Peter Lang SA, Editions scientifiques internationales, Berne 2016  
Wabernstrasse 40, CH-3007 Berne, Suisse  
[info@peterlang.com](mailto:info@peterlang.com), [www.peterlang.com](http://www.peterlang.com)

## Remerciements

Nous tenons à remercier ici Pauline Dubosson pour son aide précieuse dans le travail de mise en forme de ce volume.

Thierry Herman tient en particulier à remercier Steve Oswald de lui avoir proposé ce projet et de l'avoir coordonné de manière impeccable dans une atmosphère de travail très agréable.





## Table des matières / Contents

Remerciements .....	V
Table des matières / Contents .....	VII
Introduction .....	1
Thierry HERMAN & Steve OSWALD	
La vigilance épistémique .....	21
Dan SPERBER, Fabrice CLÉMENT, Christophe HEINTZ, Olivier MASCARO, Hugo MERCIER, Gloria ORIGGI & Deirdre WILSON	
Manipulation et cognition: un modèle pragmatique.....	69
Didier MAILLAT	
Biased argumentation and critical thinking.....	89
Vasco CORREIA	
Vers une naturalisation de la rhétorique? Problèmes épistémologiques ...	111
Emmanuelle DANBLON	
A case for emotion awareness.....	129
Evgenia PAPAROUNI	
L'argument d'autorité: de sa structure à ses effets .....	153
Thierry HERMAN	
Argumentation from expert opinion in science journalism: The case of Eureka's Fight Club.....	185
Sara GRECO MORASSO & Carlo MORASSO	
Two-sided rhetorical strategies in top management's letters to shareholders and stakeholders from corporate reports.....	215
Ioana Agatha FILIMON	
Presupposing redefinitions .....	249
Fabrizio MACAGNO	
Présuppositions discursives, assertion d'arrière-plan et persuasion.....	279
Louis de SAUSSURE	
Pragmatics, cognitive heuristics and the straw man fallacy .....	313
Steve OSWALD & Marcin LEWIŃSKI	
Résumés / Abstracts.....	345



# Introduction

Thierry HERMAN, Université de Neuchâtel & Université de Lausanne  
Steve OSWALD, Université de Fribourg & Université de Neuchâtel

D'une certaine manière, la rhétorique est un art ou une technique – plutôt qu'une science – cognitive. Beaucoup de travaux ont relevé la pertinence sociale de la rhétorique, qui se déploie à travers différents genres de discours, le délibératif, le judiciaire et l'épidictique, lesquels représentent différents lieux sociaux: l'agora, le tribunal, la Cité. Mais l'art de discourir en situation concrète dans l'espoir de faire adhérer l'auditoire à une thèse suppose une forte aptitude cognitive qu'on appellerait aujourd'hui de la méta-représentation. Les traités rhétoriques antiques, on le sait, sont littéralement innervés par cette dimension cardinale de l'art oratoire: l'adaptation à l'auditoire. Comment puis-je m'en faire une représentation? Qu'est-ce qui est susceptible de l'émouvoir ou de l'agacer? Comment arranger mes arguments pour ne pas brusquer le juge? Depuis Corax et Tisias, l'enjeu majeur est de tenter de maîtriser les ressorts cognitifs du juge; sa réussite, si on en croit l'évolution de la pensée contre la rhétorique sophistique, serait alors un indice d'un pouvoir quasi-magique de la parole.

Or, à partir de moment où agir sur les représentations d'autrui est facilité par des techniques rhétoriques ou sophistiques, la question de la tromperie verbale s'est immiscée dans des affaires de régulation sociale et, avec elle, des enjeux tant de crédibilité que de crédulité. La naissance mythique de la rhétorique fondée sur une parole libérée et des institutions démocratiques met paradoxalement en scène une forme de tyrannie de la parole par ceux qui la maîtrisent et qui exploitent notre propension naturelle à prendre des raccourcis dans le raisonnement. La psychologie, les sciences cognitives et comportementales ont largement illustré maintenant nos illusions cognitives (Pohl 2004), nos raisonnements à deux vitesses (Kahneman 2011, Evans & Over 1996) tout comme la persuasion par des voies périphériques (Petty & Cacioppo 1986) et des heuristiques approximatives (Tversky & Kahneman 1974, Gigerenzer *et al.* 2011). Et on peut supposer que les sophistes ont pris conscience de cette fragilité intrinsèque de notre art de raisonner. L'exemple

totémique du procès de Phryné, où la beauté de Phryné suffit à retourner une cause que l'on croyait perdue, illustre parfaitement la peur d'être persuadé pour de mauvaises raisons:

Hypéride, l'avocat de Phryné, n'ayant pas réussi à émouvoir les juges et se doutant qu'ils allaient la condamner, décida de la mettre bien en vue, déchira sa tunique et dévoila sa poitrine à tout le monde. A ce moment, il tint des arguments si pathétiques que les juges, pris soudain d'une frayeur superstitieuse vis-à-vis d'une servante et prêtresse d'Aphrodite, se laissèrent gagner par la pitié et s'abstinrent de la mettre à la mort. (Athénée de Naucratis, *Deipnosophistes*, XIII, 59)

Dans le cadre démocratique rendant encore plus aigu une forme de dépendance à l'information d'autrui, la nécessité de croire tout comme la possibilité de se faire avoir mettent à l'épreuve tant le fonctionnement social de la Cité que l'évaluation des informations et de leurs auteurs. La rhétorique du vraisemblable plutôt que du vrai, la rhétorique des opinions plutôt que du juste et de l'injuste (Platon, *Gorgias*, 455a) est à la fois une chance – la rhétorique est fille de la démocratie, elle en sert les principales institutions – et une menace. Cette tension illustrée par l'idéale «rhétorique pour les dieux» (Danblon 2005: 31) souhaitée par Platon contre la rhétorique à hauteur d'homme professée par Aristote est bien documentée sur le plan philosophique et historique. Mais l'éclairage jeté par les sciences cognitives dans ce débat nous paraît assez fascinant pour justifier ce volume. Le but de plusieurs contributions dans cet ouvrage n'est pas de dénoncer les effets de certains schèmes argumentatifs que d'aucuns jugeraient fallacieux ni d'ajouter une couche nouvelle aux critiques des sophismes. De Platon aux écoles d'argumentation modernes professant le *Critical Thinking* (Herman 2011), des normes critiques sont édifiées contre les abus de langage, les déviations argumentatives, les illusions de logique ou paralogismes (cf., par ex. Hamblin 1970). On peut évidemment enseigner les arguments fallacieux et comment les contrer – encore que cela pose quelques problèmes: comment distinguer le fallacieux du maladroit, l'abus de l'approximation, la rigueur argumentative de la liberté rhétorique prise avec elle? Mais les classiques mouvements réputés fallacieux comme l'*ad hominem* (Johnstone 1978, Walton 1998), l'*ad populum* (Maillat, ce volume) et l'*ad verecundiam* (Herman, ce volume) sont souvent réinterprétés comme moins sophistiques qu'il n'y paraît lorsqu'ils ne sont pas considérés comme une façon courante d'argumenter: «le paralogisme n'est pas l'exception, il est la règle» (Angenot 2008: 92). L'enjeu de ce

volume ne se situe dès lors ni dans la prophylaxie vis-à-vis des sophismes ni dans leurs remèdes, autrement dit ni dans l'amont ni dans l'aval des sophismes et paralogismes, mais dans leur fonctionnement *hic et nunc*. Quels sont les mécanismes langagiers et cognitifs qui expliquent la «performance» des arguments réputés fallacieux? Ce livre, issu du colloque *Communication et Cognition: manipulation, persuasion et biais dans le langage*, tenu à Neuchâtel du 26 au 28 janvier 2011, met clairement en perspective l'interface langue-cognition-société à travers des contributions de plusieurs auteurs qui ont été relus en double-aveugle par un collègue d'experts que nous remercions ici.

Plusieurs contributions font explicitement ou non référence à l'hypothèse d'un nécessaire filtre cognitif de vigilance épistémique: «nous ne pourrions pas faire preuve de confiance mutuelle sans vigilance mutuelle» (Sperber *et al.*, ce volume). Nous avons choisi de traduire en français cet article en tête de notre ouvrage pour fournir un cadre explicatif intéressant à bien des phénomènes observés ici. L'avantageuse posture de confiance vis-à-vis de l'information communiquée par autrui, la présomption de vérité sur laquelle la communication humaine doit pouvoir compter demande dans le même temps, dans un contexte asymétrique d'intérêts divergents, une possibilité de circonspection proche de ce que la rhétorique appelle la *phronèsis* ou la prudence. La mise à contribution voire la mise à l'épreuve de la vigilance épistémique se fait jour lorsque, par exemple, un juge se voit présenter un schème argumentatif comme le *corax* visant à plaider l'innocence d'un prévenu non à cause du manque d'indices et motifs, mais de leur surabondance suspecte. Le retournement d'un argument, qui fascinait les anciens (cf. les exercices des *dissoi logoi* ou des antilogies), voire le retournement des convictions illustré par exemple par le film *Douze homme en colère* (Lumet 1957), nous irrémédiablement les enjeux sociaux de la rhétorique aux effets cognitifs de la parole argumentée et à sa potentielle puissance persuasive. Quand bien même l'article de Sperber *et al.* ne mentionne nullement la rhétorique, les connaisseurs de cette discipline ne seront pas surpris de constater que les facteurs qui influencent l'acceptation ou le rejet d'une information relèvent de sa source ou de son contenu. Pour la première, les auteurs estiment que la fiabilité d'une source se mesure à l'aune de sa compétence et de sa bienveillance; on n'est pas très loin de l'*ethos* aristotélicien composé de prudence (*phronèsis*), vertu (*arête*) et bienveillance (*eunoïa*). Pour la seconde, les auteurs mobilisent la théorie de la pertinence

pour expliquer le mécanisme d'évaluation des informations et la mobilisation des informations contextuelles d'arrière-plan, mobilisant ainsi un rapport entre sciences du langage et cognition, alors que la dimension sociale est interrogée en fin d'article par une transposition de la vigilance épistémique à l'échelle de la population.

Ce cadre théorique stimulant et innovant est exploité par Didier Maillat pour traiter le cas de la manipulation (voir aussi Maillat & Oswald 2009, 2011). Ne la considérant pas du point de vue de l'émetteur ou de ses intentions, mais du point de vue des destinataires, Maillat envisage la manipulation comme un effet sur les processus interprétatifs mobilisés. Elle s'exerce par des contraintes de sélection d'un contexte pertinent: soit on rend plus accessible un contexte dans lequel l'énoncé manipulateur se voit renforcé, soit on affaiblit le contexte dans lequel l'énoncé manipulateur serait inconsistant et éliminé. En somme, la manipulation profite d'une propension à rechercher l'information la plus pertinente à moindre coût: au cas où un contexte est pertinent pour rejeter l'énoncé E, il faudra faire en sorte que l'effort pour accéder à ce contexte soit plus important que celui nécessaire pour mobiliser un contexte dans lequel E est admis. Maillat illustre cela par le matraquage des arguments dans la publicité et la propagande pour augmenter le degré d'accessibilité d'un slogan par exemple. La dimension langagière de cette exploration est illustrée par l'*ad populum*, mouvement simulant sur le plan langagier l'effet de la répétition physique d'une même idée par des sources différentes. Sur le plan rhétorique, Maillat illustre par son article un des aspects de la question cruciale de l'efficacité discursive. Cette quête éternelle de la *technè* rhétorique pour les procédés les plus efficaces dans la construction du discours – *inventio*, *dispositio* et *elocutio* – trouve ici, du côté de la réception et des effets cognitifs du destinataire, une hypothèse explicative sur l'efficacité de certains procédés trompeurs. Cette proposition illustre du même coup le potentiel dialogue entre rhétorique et cognition que nous souhaitons mettre en place dans cet ouvrage et le colloque dont il est issu.

La contribution de Vasco Correia traite frontalement de la problématique du lien entre biais cognitifs et arguments fallacieux dans une perspective non de réception, mais de production d'arguments. Il défend en particulier l'idée que le raisonnement soutenu par une motivation biaisée est responsable des erreurs argumentatives non intentionnelles que les sujets parlants peuvent

être amenés à commettre dans les échanges argumentatifs. Correia présente trois biais motivationnels liés aux émotions, à savoir prendre ses désirs pour des réalités (*wishful thinking*), croire  $p$  par peur de  $non-p$  (*aversive thinking*) et prendre ses inquiétudes pour des réalités (*fretful thinking*), comme lorsque, sans nouvelle de son enfant depuis des heures, on pense qu'il lui est arrivé malheur; il souligne dans un deuxième temps les liens entre ces trois biais et des formes reconnues d'argumentation fallacieuse – l'argument de la pente glissante pour le *fretful thinking* par exemple, puis discute de différentes techniques destinées à contrecarrer l'effet de ces biais pour assurer la qualité d'un échange argumentatif. Correia prend résolument le parti de s'intéresser à la production d'arguments, dans une perspective dialogique, afin de montrer que le caractère fallacieux de certains arguments peut s'expliquer en faisant intervenir un modèle cognitif du traitement de l'information, notamment en ce qui concerne les motivations sous-jacentes aux croyances. Il envisage en fin de parcours différentes techniques d'autocontrôle des biais motivationnels à même d'assurer une pratique raisonnée, et même éthique, d'un idéal argumentatif. En ceci, Correia pourrait se montrer plus platonicien qu'aristotélicien, quand bien même sa contribution est majoritairement soutenue par une volonté explicative et non normative. Il illustre d'une certaine manière que les locuteurs les mieux intentionnés du monde peuvent produire des arguments fragiles sinon erronés qu'aucune norme argumentative ou de raisonnement ne peut saisir, d'autant moins que ses locuteurs peuvent rationaliser *a posteriori* leurs croyances acquises sous l'influence des trois biais motivationnels décrits. Or, en ceci, il ramène la rhétorique à hauteur d'homme et rappelle que la faiblesse argumentative n'est pas qu'une faiblesse de raisonnement, mais aussi une faiblesse humaine.

Emmanuelle Danblon fait écho à plusieurs contributions déjà mentionnées en s'inscrivant dans un cadre radicalement naturaliste et interdisciplinaire cherchant tant à éviter le réductionnisme que le relativisme post-moderne. Le cadre naturaliste est celui largement adopté par les sciences cognitives et la transversalité de la rhétorique est une préoccupation continue de cet ouvrage. Emettant des réserves par rapport à la vision naturaliste de Sperber, considérée par elle comme réductionniste, Danblon insiste sur la dimension pratique de l'intelligence humaine. Conduisant son enquête à partir de la célèbre dichotomie entre convaincre et persuader qui mène Chaim Perelman et Lucie Olbrechts-Tyteca à assouplir leur notion de

rationalité, Danblon propose de penser la persuasion hors de la question limitée des effets, mais dans un cadre technique au sens propre de *tekhnè*, à savoir l'artisanat. Dans ce cadre, la *praxis*, fait de règles, mais aussi d'expériences, d'intuition, le tout au sein de cadres sociaux, permet la production rhétorique, ou *poiésis*, visant l'efficacité pour atteindre une visée, *poiésis* rendue possible par, et qui rend possible, des dispositions citoyennes à persuader comme à être persuadé en retour, l'*hèxis*. Danblon inscrit dès lors sa réflexion dans un cadre naturaliste sans se cantonner à une «dimension purement cognitive de la raison»; le cadre social, concrètement pragmatique et citoyen de la persuasion autorise le développement d'une forme d'agilité artisanale qu'est l'exercice de la rhétorique, en dehors de toute considération d'ordre moral.

Evgenia Paparouni discute du rôle joué par les émotions dans l'analyse de l'argumentation, en particulier depuis un point de vue rhétorique. Le problème principal discuté est le suivant: si d'une part l'émotion est traditionnellement considérée comme un obstacle à la discussion raisonnée, et si d'autre part l'émotion, en rhétorique, est toutefois considérée comme un moyen de conviction/persuasion au même titre que l'*ethos* et le *logos*, comment envisager l'émotion dans un cadre rhétorique de recherche cohérent (c'est-à-dire qui puisse évacuer le problème du caractère fallacieux traditionnellement imputé aux émotions tout en posant des limites à leur utilisation)? Cette contribution examine les apports des sciences cognitives dans le domaine de l'émotion ainsi que ceux de la philosophie de l'esprit, pour montrer que le dualisme cartésien n'est plus de mise dans le domaine et relever que cela permet dès lors de réfléchir sur la compatibilité entre une approche rhétorique et une approche cognitive. La question à laquelle la pratique rhétorique peut aider à répondre n'est donc plus celle de savoir si les émotions sont responsables des dérives de l'argumentation (en tant qu'elles seraient instrumentales à la manipulation), mais plutôt celle de savoir comment la pratique de la rhétorique – avec sa composante émotive – peut constituer un moyen responsable d'éviter une telle dérive.

Thierry Herman propose pour sa part de revenir sur une autre ressource rhétorique fréquemment considérée comme fallacieuse: l'appel à l'autorité. Ancrée dans une dimension linguistique et rhétorique, sa réflexion intègre une dimension cognitive dans la mesure où il enquête sur les effets du lien entre l'assertion sur le plan langagier et l'autorité sur le plan rhétorique. Il



montre que l'argumentation d'autorité ne saurait se réduire à une description classique de l'argumentation et impose moins une vérité du monde que l'idée d'indubitabilité de cette supposée vérité. Constatant que l'argument d'autorité se réduit empiriquement le plus souvent à une prémisse – la conclusion *X est vraie* étant triviale – Herman montre les effets d'une argumentation d'autorité qui se voit greffée à une argumentation plus complète, ce qui favorise entre autres l'idée d'imposer une certitude. Allant jusqu'au bout d'un travail de désossement de l'argument d'autorité, Herman évalue ses différents critères définitoires et ose un rapprochement entre l'argument d'autorité et l'assertion autoritaire, les deux partageant l'idée d'imposer un contenu comme certain. Il étudie à ce titre différents exemples de formes langagières et les effets d'une forme de rhétorique de l'autorité.

Dans une forme de continuité avec la réflexion précédente, Sara Greco Morasso et Carlo Morasso investiguent le cas particulier de l'argument de l'autorité experte dans le cadre du journalisme scientifique. L'appel à l'autorité est vu dans une tension entre la nécessité de faire confiance à l'autorité scientifique et le scepticisme vis-à-vis de celle-ci dans le cadre d'une relation asymétrique entre expert et public. L'étude sur l'appel à l'autorité dans le journalisme scientifique – alors même qu'il est mal considéré par les scientifiques eux-mêmes – s'appuie sur un exemple analysé dans le cadre de deux principaux modèles d'analyse de l'argumentation: les questions critiques de Walton et le modèle topique tessinois (AMT). Elle montre qu'une ressource fallacieuse de l'autorité est de rester vague sur les sources mobilisées. Le caractère fallacieux ou non de l'argument d'autorité est ainsi revisité, jusqu'à proposer une nouvelle question critique à la liste de Douglas Walton.

La contribution de Ioana Agatha Filimon étudie les caractéristiques et l'emploi de messages ambivalents (définis comme des messages qui défendent une conclusion, mais qui mentionnent également une conclusion ou des contre-arguments adverses) dans un corpus très spécifique: les lettres d'introduction écrites par les instances dirigeantes d'une entreprise, destinées soit aux actionnaires soit aux parties prenantes (partenaires, etc.). Filimon combine de manière originale les outils de la théorie de l'argumentation (la théorie pragma-dialectique en particulier) et ceux de la recherche sur la persuasion (d'après O'Keefe principalement). Cette contribution sur le plan rhétorique a des implications sur celui de la cognition. D'une part, ce chapitre

tente d'apporter une réponse aux deux questions majeures de l'argumentation (sur la base de l'étude d'un corpus donné): les arguments sont-ils valides/acceptables? Les arguments sont-ils efficaces? D'autre part, l'analyse de Filimon permet de dégager les contraintes liées au genre du corpus, mais aussi des contraintes pesant sur la crédibilité de tels messages: ses résultats montrent que les instances dirigeantes des entreprises sont très souvent amenées à faire usage de messages ambivalents dans le but d'augmenter leur crédibilité.

Fabrizio Macagno étudie le schème particulier de l'argument par définition. A l'instar d'Herman qui montre ici les effets de l'autorité présumée, Macagno s'intéresse à la désignation de la réalité lorsque la définition de cette réalité est laissée implicite, considérée comme acquise. Cet acte de présumer une redéfinition est étudié sous l'angle des effets persuasifs possibles. S'appuyant sur les travaux de Stevenson, Macagno rappelle ainsi la composante descriptive mais aussi émotionnelle des définitions et des désignations, poussant par exemple l'administration Obama à ne plus utiliser le syntagme «guerre contre le terrorisme», mais «*Overseas Contingency Operations*». Etudiant les différentes logiques profondes qui structurent les arguments de classification, lesquelles s'occupent plus de la part descriptive de la désignation que de la part émotionnelle, Macagno inscrit sa réflexion dans le cadre de schèmes liés à des valeurs, qui sont susceptibles de provoquer l'action. Plus largement, il cherche à montrer comment les mots redéfinis – Obama impliquant par exemple que parler d'hostilités en Libye est impossible tant que des troupes ne se battent pas sur le terrain – peuvent provoquer l'action et rappelle l'importance de la présupposition linguistique qu'il considère sous la forme d'un raisonnement de présomption. Le pouvoir persuasif et manipulateur de s'appuyer sur une redéfinition présumée et *a priori* non-partagée plutôt qu'exposée est ainsi mis en lumière. Il s'inscrit dans la ligne directrice de cet ouvrage: les effets cognitifs de stratégies linguistiques dans des contextes rhétoriques de tentatives de persuasion.

Avec la contribution de Louis de Saussure, la réflexion sur l'importance de la présupposition dans les mouvements persuasifs est prolongée. Examinant les questions d'accommodation des présuppositions, Saussure considère le phénomène de présupposition de manière large en ce que des contenus inférés peuvent aussi être considérés comme des présuppositions.

A mi-chemin entre la présupposition stricte et les implicatures, les présuppositions qu'il appelle discursives demandent à être inférées pour faire sens des intentions du locuteur. Dire «Les armes sont interdites dans ce secteur» conduit à tirer des inférences du type «Les armes peuvent être autorisées dans d'autres secteurs» qui ne sont ni présupposées, ni conversationnellement implicites. Souvent, elles font partie de l'arrière-plan conversationnel, mais il peut arriver que ce ne soit pas le cas: en cela, elles peuvent aussi être accommodées, faire faussement croire à de l'information ancienne. Ainsi, le vote suisse sur l'interdiction des minarets présupposait discursivement une forme de dangerosité ou de menace pour que les citoyens puissent faire sens de cet objet de vote. Cette présupposition discursive est hautement pertinente et pourtant se situe en arrière-plan: Saussure rend compte d'une pertinence d'arrière-plan. Cette contribution complète ainsi plusieurs autres propositions de ce volume en montrant les effets de persuasion qui semblent échapper à une structure argumentative explicite et limiter notre vigilance épistémique pour faire admettre une vision du monde.

Le chapitre d'Oswald & Lewiński clôt l'ouvrage et reprend la ligne directrice de celui-ci en proposant une analyse des mécanismes cognitifs sur lesquels le sophisme bien connu de l'homme de paille s'appuie, en proposant une explication à l'interface de la rhétorique et des sciences cognitives. Les auteurs discutent des liens entre heuristiques cognitives et argumentation fallacieuse à la lumière de la notion de vigilance épistémique proposée par Sperber *et al* en début de volume. Leur prolongement de cette réflexion amène Oswald et Lewiński à considérer que si les arguments fallacieux ont des chances de couronner de succès une entreprise de persuasion, c'est précisément parce que d'un point de vue cognitif ils agissent sur nos filtres de vigilance épistémique; en ce sens, le sophisme génère une illusion de compétence et de bienveillance (quant à la source d'information) dans le même temps qu'il mime la cohérence d'un propos. C'est donc en évacuant toute réticence critique – ou en l'empêchant – qu'un sophisme est victorieux. Le cas de l'homme de paille, argument fallacieux tablant sur la mésattribution d'un propos à un locuteur, est discuté à la lumière d'un exemple issu d'un débat politique lors d'une séance du Conseil National suisse. L'analyse proposée montre en particulier que les chances de succès de l'homme de paille reposent en grande partie sur sa capacité de rendre la mésattribution

sur laquelle il repose indétectable, soit, en termes cognitifs, non pertinente (difficilement accessible et épistémiquement faible).

Au final, même si certaines contributions diffèrent profondément, nous avons voulu un livre traversé d'une part par l'interdisciplinarité nécessaire entre sciences du langage, sciences cognitives et sciences sociales qu'exige la rhétorique et, d'autre part, par l'examen de multiples schèmes argumentatifs, fallacieux ou non, dans leurs formes, leurs effets et, pour certaines contributions, dans le cadre de pratiques discursives délimitées. Adoptant une posture descriptive et explicative visant à comprendre la mécanique de ces schèmes, les contributeurs de ce volume montrent, explicitement ou non, l'avantage de reconsidérer la rhétorique sous l'angle cognitif, avantage qui nous semblait assez peu mis en valeur dans les publications récentes révélant, elles, toute la vivacité et l'actualité de la rhétorique. Plusieurs propositions originales ou hypothèses stimulantes émaillent cet ouvrage: on espère qu'elles inspireront tant les chercheurs spécialisés en rhétorique et sciences du langage à aller voir du côté de la psychologie cognitive que les spécialistes de ce domaine à mettre en évidence la rhétoricité de leurs recherches.

## Introduction (English version)

In a way, rhetoric is an art or a cognitive technique – more than a science. Numerous works have highlighted the social relevance of rhetoric, which spreads across different discourse genres, such as the deliberative, the judicial and the epideictic. These in turn represent different social links: the *agora*, the court of law, the city. Yet, the art of expressing oneself in concrete situations in the hope of gaining the audience's consent on a given issue requires the operation of a cognitive ability that is nowadays referred to with the term *metarepresentation*. Ancient treatises of rhetoric, as we know, abound with this cardinal dimension of oratory art, namely audience adaptation. How can I achieve an accurate representation of the audience's values, beliefs and other mental states? What is likely to move or to annoy its members? How can I design my arguments in order not to rush the judge? Ever since Corax and Tisias, the major stake in this endeavour has been to master the cognitive responses of the judge; if we follow the evolution of thought that gradually built against sophistic rhetoric over the centuries, it appears that the success of the latter is an indication of the quasi-magical power of words.

Nonetheless, once we consider that rhetorical or sophistic techniques influence people's representations, verbal deception becomes a matter of social regulation, together with issues of credibility and credulity. The mythical birth of rhetoric, which was founded on the possibility of unconstrained speech and the emergence of democratic institutions, thus paradoxically instantiates a form of speech tyranny reserved to those who master it and are able to exploit our natural inclination to take reasoning shortcuts. Cognitive and behavioural sciences have largely documented and illustrated cognitive illusions (Pohl 2004), dual-system inferences (Kahneman 2011, Evans & Over 1996), peripheral routes to persuasion (Petty & Cacioppo 1986) and approximate heuristics (Tversky & Kahneman 1974, Gigerenzer *et al.* 2011). And yet, we could surmise that sophists already had some awareness of the intrinsic fragility of our art of reasoning. The distinctive example of Phryne's trial, where the beauty of Phryne managed to win an apparently lost case, perfectly illustrates the fear of being persuaded for the wrong reasons:

Failing to move the judges and anticipating her condemnation, Hyperides, Phryne's advocate, decided to expose her by tearing her tunic apart, thus revealing her chest to everyone. He then formulated arguments that were so pathetic that the judges, suddenly falling prey to the superstitious fear of a servant and priestess of Aphrodite, were swayed by mercy and refrained from putting her to death. (Athenaeus of Naucratis, *Deipnosophistae*, XIII, 59)

In a democratic context fostering a form of dependence towards other people's information, the necessity of believing everything and the possibility of being duped are challenges for both the social management of the City and the evaluation of information (and of its source). The rhetoric of plausibility – more than the rhetoric of truth – and the rhetoric of opinion – more than that of the fair and the unfair (Plato, *Gorgias*, 455a) – are at the same time a blessing, insofar as rhetoric is the daughter of democracy and serves its main institutions, and a threat. The tension between the ideal “rhetoric for Gods” (Danblon 2005: 31) Plato had wished for and Aristotle's down-to-earth rhetoric has philosophically and historically been well documented; we believe the insights cognitive science is able to contribute to this debate justify this volume. However, the contribution of several chapters of this book is neither to be found in the condemnation of the fallacious effects of specific argument schemes nor in the addition of yet another layer to fallacy criticism. From Plato to modern schools of argumentation professing Critical Thinking (Herman 2011), critical norms have been summoned against language abuse, argumentative fault, logical illusions and so forth (see e.g. Hamblin 1970). We can of course teach fallacious arguments and ways to fight them – but this is still problematic: how do we distinguish fallaciousness from clumsiness, abuse from approximation, argumentative rigour from the rhetorical freedom it goes with? Yet, some argumentative movements traditionally classified as fallacious, such as the *ad hominem* (Johnstone 1978, Walton 1998), the *ad populum* (Maillat, this volume) and the *ad verecundiam* (Herman, this volume), are oftentimes reinterpreted as less fallacious than they seem, if not as common argumentative resources: “the fallacy is not the exception, it is the rule” (Angenot 2008: 92). This volume therefore proposes neither some sort of prophylaxis against fallacies, nor tentative remedies to their ‘nuisance’; in a nutshell, this volume is not about what happens in communication before or after a fallacy is committed, but rather about the way they work *hic et nunc*, and thus about the cognitive and linguistic mechanisms at play when

fallacious arguments are ‘performed’. This book, which gathers papers presented during the international conference *Communication & Cognition: manipulation, persuasion and biases in language*, held at the University of Neuchâtel in January 2011, explores the interface between language, cognition and society, through contributions which have been double-blind reviewed by a panel of experts, whom we take the opportunity to thank here.

Several contributions explicitly or implicitly take on board the assumption that we necessarily have some cognitive filter of epistemic vigilance: “We could not be mutually trustful *unless* we were mutually vigilant” (Sperber *et al.* 2010: 364). We have thus decided to open our volume with a French translation of “Epistemic vigilance” to provide a stimulating explanatory framework for many of the phenomena discussed throughout the book. The advantageous stance of trust towards communicated information and the presumption of truth on which human communication must be able to rely require at the same time, in asymmetrical contexts of diverging interests, the possibility of circumspection in the vein of what rhetoric calls *phronêsis*, or cautiousness. When a judge is presented with an argumentative scheme such as the *corax*, which defends the innocence of a suspect with the suspicious abundance of clues and motives rather than with the lack thereof, is to us a typical situation in which the possibilities offered by the framework of epistemic vigilance can be exploited, and even tested. The possibility of overturning arguments (which fascinated the classical tradition, as evidenced by *dissoi logoi* exercises and antilogies) and convictions (as illustrated for instance in the film *Twelve angry men* (Lumet 1957)) irremediably ties the social stakes of rhetoric to the cognitive and potentially persuasive effects of argumentative speech. Even if Sperber *et al.* make no mention of rhetoric, scholars in the discipline will not be surprised to find that the factors influencing acceptance or rejection of a given piece of information have to do with its source and its content. The authors consider that source reliability can be measured in terms of competence and benevolence: this is reminiscent of the Aristotelian *ethos*, made of prudence (*phronêsis*), virtue (*arête*) and benevolence (*eunoia*). In terms of content, the authors draw on Relevance Theory to account for the mechanisms information evaluation and background assumption selection. In so doing, they explore the relationship between cognition and the language sciences, while the social

dimension of the phenomenon is tackled towards the end of the paper by transposing epistemic vigilance on the scale of a population.

Didier Maillat exploits this stimulating and innovative framework to account for manipulation (see also Maillat & Oswald 2009, 2011). Concerned with reception rather than production, Maillat construes manipulation as a constraint on interpretative procedures. Manipulation enforces selection constraints on relevant contexts: either it makes accessible a context in which the manipulative utterance is found relevant or it weakens a context in which the manipulative utterance would be found inconsistent and eliminated. In a nutshell, manipulation takes advantage of our natural propensity to look for relevant information in cost-effective ways: in case a context is relevant enough to reject the manipulative utterance, the manipulator will need to make the effort required to access that context greater than the effort required to mobilise a context in which the target utterance is unchallenged. Maillat illustrates his point with the relentlessness with which some arguments in advertising and propaganda are repeated, in order to increase the accessibility of a given slogan, for instance. The linguistic dimension of such strategies is illustrated by the *ad populum* fallacy, which according to Maillat is an argumentative movement that linguistically simulates the effects of physical repetition of a same idea by different sources. From a rhetorical viewpoint, Maillat thus explores one of the crucial aspects of discursive effectiveness. The eternal quest for rhetorical *technè* in effective strategies of discursive construction – *invention*, *dispositio* and *elocution* – accordingly finds in his contribution an explanatory assumption regarding the effectiveness of certain fallacious moves. At the same time, his proposal makes room for a potential dialogue between rhetoric and cognition, which is one of the main concerns behind this volume and the conference from which it originated.

Vasco Correia's contribution frontally addresses the relationship between cognitive biases and fallacious arguments, this time not from the perspective of argument reception, but from that of argument production. He specifically defends the idea that biased motivations are responsible for the unintentional argumentative errors that people can be led to commit. Correia presents three motivational biases linked to emotions, namely wishful thinking, aversive thinking and fretful thinking. He then highlights the links between these three biases and recognized patterns of fallacious argumentation (the slippery slope for fretful thinking, for instance) and discusses different



techniques meant to counter the effect of these biases to ensure the quality of argumentative exchanges. Correia accordingly addresses argument production in a dialogical perspective to show that the fallacious nature of certain arguments can be accounted for by a cognitive model of information processing, notably in terms of the motivations at play behind beliefs. Towards the end of the chapter, he reflects on different self-control techniques meant to constrain these motivational biases, in an attempt to ensure a reasoned, and even ethical, ideal of argumentative practice. Correia could therefore be seen as being closer to Plato than to Aristotle, despite his contribution being largely characterized by an explanatory rather than normative outlook. He illustrates in a way that well-intended speakers can commit fragile or fallacious arguments that no argumentative or reasoning norm could grasp, even more so if we consider that speakers may always *a posteriori* rationalize the beliefs they have acquired under the influence of motivational biases. By the same token, he brings rhetoric back to the measure of man and reminds us that argumentative weakness is not only the weakness of the reason, but also the weakness of man.

Emmanuelle Danblon echoes these contributions by situating her work in a radically naturalistic and interdisciplinary framework meant to avoid both reductionism and post-modern relativism. While formulating some reservations towards Sperber's naturalistic vision, which she considers to be reductionist, Danblon insists on the practical dimension of human intelligence. Following Perelman & Olbrechts-Tyteca's famous dichotomy between convincing and persuading, which loosens their notion of rationality, Danblon reflects on persuasion outside the boundaries of its effects in a technical framework compatible with the proper sense of *technè*, i.e., craftsmanship. Within that framework, the *praxis*, constituted of rules, experiences and intuitions throughout many social frames, allows for rhetorical production, or *poiésis*, which targets effectiveness in view of reaching a goal. The *poiésis* is in turn made possible by – and also makes possible – both the civic disposition to persuade and that of being persuaded, i.e. the *hèxis*. As a consequence, Danblon speaks from a naturalistic stance and tries to avoid the “purely cognitive dimension of reason”. The social context of persuasion, which is civic and pragmatic in nature, licenses the development of rhetoric, which we could then define as a form of ‘handcrafted dexterity’ in the absence of any moral considerations.

Evgenia Paparouni discusses the role played by emotions in the analysis of argumentation from a rhetorical perspective. The main issue she tackles is the following: if, on the one hand, emotion is traditionally considered to be an obstacle to reasoned discussion, and if, on the other, emotion is nevertheless considered to be a rhetorical means of conviction/persuasion on a par with *ethos* and *logos*, how can emotion partake in a coherent research framework for rhetoric (i.e., a framework that could dispose of the traditional problem of linking fallacious argumentation to the use of emotions, and at the same time a framework that could place limits upon their use)? This contribution examines the input cognitive science and the philosophy of mind can provide to the study of emotions, in order to show that Cartesian dualism is no longer an option in the field and that this entails that we are now free to think about the compatibility between rhetoric and cognitive science. The question that rhetorical practice might help answering is no longer that of knowing if emotions are responsible for argumentative flaws (to the extent that they are instrumental to manipulation), but rather that of knowing how the practice of rhetoric – with an emotional component – can constitute a responsible way of avoiding such flaws.

Thierry Herman sets out to rethink another rhetorical resource traditionally considered to be fallacious, namely the appeal to authority. Grounded in a linguistic and rhetorical framework, his reflection integrates a cognitive dimension as he investigates the effects of the relationship between the linguistic act of asserting and the rhetorical status of authority. He shows that arguments from authority cannot be reduced to their classical description and that they impose the idea of indubitability of a supposed truth, rather than the factual truth itself. While observing that the argument from authority is usually realised in discourse as a single premise – the conclusion *X is true* being trivial –, Herman shows how an argument from authority partakes in a more complex argumentation, which, among other things, favours the idea of imposing a certainty. As he further decomposes the structure of the argument from authority, Herman evaluates the different definitional criteria he comes up with and proposes to treat the argument from authority and authoritative assertions on a par, as both these phenomena share the idea of imposing the certainty of a given content. He investigates different linguistic structures in this framework and the effects of this rhetorical form of authority.

Somehow in continuity with the previous contribution, Sara Greco Morasso and Carlo Morasso address the particular case of arguments from expert opinion in scientific journalism. The appeal to authority is seen here in the tension between the necessity of trusting scientific expertise and the scepticism towards it within an asymmetrical relationship between experts and the general public. This study of appeals to expert opinion in scientific journalism – even if this form of argument is itself condemned by scientists – builds on the analysis of an example according to two mainstream models of argument analysis: Walton’s critical questions model and the Argumentum Model of Topics (AMT). The results show that the fallacious resource linked to expert opinion consists in remaining vague on the sources that are mobilised to support the claim. The (non)fallacious nature of the argument is thus revisited and a novel critical question is added to Douglas Walton’s existing list of critical questions associated to the scheme.

Ioana Agatha Filimon studies the features and the uses of ambivalent messages (defined as messages that defend a claim but also contain counter-claims and counter-arguments, which are then refuted) in a very specific corpus: introductory letters to shareholders and stakeholders from annual and corporate social responsibility reports. Filimon combines in an original way the tools of argumentation theory (grounded on Pragma-Dialectics) and those provided by persuasion research (following O’Keefe for the most part). This contribution is resolutely situated in the study of rhetoric but has strong implications for cognitive research as well. On the one hand, it tries to provide an answer to the two major questions behind argumentation research (building on the analysis of a corpus): are the arguments valid/acceptable? Are the arguments effective? On the other hand, Filimon’s analysis identifies not only the constraints that are inherent to the type of corpus considered, but also those that affect the credibility of these messages: her analysis shows that the authors of these reports are very often led to use ambivalent messages in order to increase the credibility of managers.

Fabrizio Macagno’s contribution deals with the particular scheme of argument by definition. Like Herman, who addresses the effects of presupposed authority, Macagno studies the implicit redefinition of reality, which then turns out to be considered as given. The act of presupposing a redefinition is approached through the lens of its potential persuasive effects. Building on Stevenson’s work, Macagno reminds us of the descriptive

component of definitions and designations, but also of their emotional component, as he recalls how the Obama administration at one point switched from using the expression “war on terror” to privileging “Overseas Contingency Operation”. Through the study of the various fundamental logics giving their structure to arguments from classification, which cater more for the descriptive than the emotional component of designation, Macagno works his way in the framework of schemes relating to values, which are likely to prompt for action. From a more general viewpoint, he shows how redefined words – an example of which is Obama’s decision not to talk about “hostilities” in Libya as long as no troops are fighting on the field – can cause action and reminds us of the importance of linguistic presupposition, which he puts on a par with reasoning from presumption. The persuasive and deceptive power of presupposed and *a priori* unshared, rather than exposed, redefinitions is brought to the fore. His contribution is thus particularly relevant to this volume, as it tackles the cognitive effects of linguistic strategies in the rhetorical context of persuasive attempts.

With Saussure’s chapter, the reflection on the importance of presupposition in persuasive movements is given additional attention. As he examines issues of presupposition accommodation, Saussure considers the type of inferred contents covered by the notion. Somewhere in-between semantic presupposition and implicature, Saussure postulates the existence of *discursive presuppositions*, which also need to be inferred to make sense of speaker intention. Processing an utterance such as “Firearms are prohibited in this sector” triggers the inference of additional contents such as “Firearms can be authorised in other sectors”, which are neither presupposed nor conversationally implicated. They often belong to the conversational background, but this need not be the case: in this sense, they can also be accommodated and pass for old information. To take an example, the Swiss vote against the construction of minarets that took place in 2009 had to discursively presuppose some sort of danger or threat in order for voters to be able to make sense of initiative. Such a discursive presupposition is highly relevant despite being backgrounded: Saussure thus speaks of *background relevance*. This contribution resonates with many others in the volume by addressing persuasive effects beyond explicit argumentative structures, which limit or circumvent our epistemic vigilance filters to favour some vision of the world.

Oswald and Lewiński's chapter concludes our volume and specifically engages with its guiding thread by proposing an analysis of the cognitive mechanisms underlying the well-known straw man fallacy. Theirs is thus an account located at the interface of rhetoric and cognitive science. They discuss the relationship between cognitive heuristics and fallacious argumentation (yet, in complementarity to Correia's contribution, they tackle reception more than production) in light of the notion of epistemic vigilance developed by Sperber *et al.* at the beginning of the volume. Their reflection leads them to conclude that if fallacious arguments have some prospect of being effective in a persuasive attempt, it is precisely because from a cognitive perspective they influence our epistemic vigilance filters: in this sense, a fallacy generates an illusion of competence and benevolence (in terms of the source of information) and at the same time mimics the content's coherence (including its contextual relevance). It is thus by disposing of critical reservations – or by preventing them – that fallacies come to be successful. The case of the straw man fallacy, an argumentative movement involving the misattribution of content to its target, is discussed in light of an example taken from a political debate in the Swiss National Council. The analysis carried out holds in particular that the chances of success of the straw man largely rest on its capacity to make said misattribution undetectable, which, in cognitive terms, amounts to the misattribution's accessibility and epistemic strength being lowered.

In the end, even if the set of contributions to this volume is at times heterogeneous in terms of their scope, we trust that our interdisciplinary concern has allowed us to explore the interface between language science and social and cognitive science, which in our view the study of rhetoric should systematically incorporate. On the more practical side, the examination of multiple argumentative schemes, be they fallacious or not in their form, their effects and, in some contributions, in delineated discursive practices, has allowed us to explore some operational facets of rhetorical practices. By adopting a descriptive and explanatory stance tailored to understand the mechanics of these schemes, our contributors have explicitly – and sometimes implicitly – displayed the advantages of a cognitive take on rhetoric; this is a direction that has seldom been explored in recent publications in the field, which are more inclined to assess rhetoric's vivacity in contemporary contexts. A number of original proposals and stimulating

hypotheses emerge throughout this volume: we hope that these will inspire researchers in the language sciences who specialise in rhetoric to take on board cognitive scientific insights and also researchers in cognitive science to engage with the rhetoricity of their own research.

## Bibliographie / References

- Angenot, M. (2008): *Dialogues de sourds*. Mille et Une Nuits, Paris.
- Danblon, E. (2005): *La fonction persuasive*. Armand Colin, Paris.
- Evans, J. S. B. T. & Over, D. E. (1996): *Rationality and Reasoning*. Psychology Press, Hove.
- Gigerenzer, G., Hertwig, R. & Pachur, T. (eds) (2011): *Heuristics: The foundation of adaptive behavior*. Oxford University Press, New York.
- Hamblin, Ch. (1970): *Fallacies*. Methuen, London.
- Herman, T. (1996): «Le courant du Critical Thinking et l'évidence des normes: réflexions pour une analyse critique de l'argumentation», *A contrario*, 16, 41-62.
- Johnstone, H. W. Jr. (1978): *Validity and Rhetoric in Philosophical Argument*. The Dialogue Press of Man and World, University Park PA.
- Kahneman, D., (2011): *Thinking, Fast and Slow*. Penguin Group, London.
- Maillat, D. & Oswald, S. (2009): «Defining manipulative discourse: the pragmatics of cognitive illusions», *International Review of Pragmatics* 1(2), 348-370.
- (2011): «Constraining context: a pragmatic account of cognitive manipulation», in Hart, C. (éd.), *Critical discourse studies in context and cognition*. John Benjamins, Amsterdam, 65-80.
- Petty, R. E. & Cacioppo, J. T. (1986): «The elaboration likelihood model of persuasion», in Berkowitz, L. (ed.), *Advances in Experimental Social Psychology*. Academic Press, Orlando, FL, 124-206.
- Pohl, R. (éd.), (2004): *Cognitive Illusions: a handbook on fallacies and biases in thinking, judgment and memory*. Psychology Press, New York.
- Tversky, A. & Kahneman, D. (1974): «Judgment under uncertainty: Heuristics and biases», *Science* 185 (4157), 1124-1131.
- Walton, D. (1998): *Ad hominem arguments*. University of Alabama Press, Tuscaloosa.

# La vigilance épistémique<sup>1</sup>

Dan SPERBER, Fabrice CLÉMENT, Christophe HEINTZ, Olivier MASCARO,  
Hugo MERCIER, Gloria ORIGGI et Deirdre WILSON

## 1. Introduction

Les humains disposent d'une série de mécanismes cognitifs de vigilance épistémique qui ciblent le risque d'être induit en erreur par autrui. Telle est la thèse que nous présentons ici en considérant les différentes manières dont la vigilance épistémique fonctionne dans la vie mentale et sociale des humains. Notre but est d'intégrer au sein d'un cadre de recherche cohérent un large éventail d'hypothèses développées ailleurs par nous-mêmes ou par d'autres, plutôt que de présenter des arguments détaillés pour chacune de ces hypothèses.

Les humains sont des animaux exceptionnels à la fois par la richesse et la puissance de leurs capacités cognitives et par le degré auquel ils s'appuient sur une large variété d'informations communiquées par autrui. Ces deux traits sont liés. D'une part, il ne serait pas possible de se reposer autant sur la richesse de la communication en l'absence de capacités cognitives spécifiques à l'espèce, en particulier le langage et une forme avancée de lecture de l'esprit (*mindreading*). D'autre part, ces capacités individuelles ne se développeraient pas ou ne fonctionneraient pas correctement en l'absence de compétences cognitives, d'outils conceptuels et de connaissances d'arrière-plan acquis par le biais d'autrui.

Jusqu'à quel point les autres sont-ils des sources d'informations fiables? En général, les autres ne se trompent pas plus souvent que nous-mêmes – après tout, 'nous' et 'eux' désignent ici les mêmes personnes – et ils savent

1 Traduction française assurée par Steve Oswald, qui remercie chaleureusement Fabrice Clément, Thierry Herman, Hugo Mercier et Dan Sperber pour leur précieuse collaboration. L'article original a été publié en 2010 dans *Mind & Language*, Vol. 25 (4), pp. 359–393, chez Blackwell qui en a aimablement autorisé la traduction.

des choses que nous ignorons. Il devrait par conséquent être avantageux de s'appuyer, même aveuglément, sur la compétence des autres. Serait-il plus avantageux de moduler notre confiance en exerçant un certain degré de vigilance vis-à-vis de leur compétence? Cela dépend du coût et de la fiabilité d'une telle vigilance. De toute façon, le problème majeur que pose l'information communiquée n'est pas lié à la compétence des autres, mais à leurs intérêts et à leur honnêteté. Si les intérêts des autres se confondent souvent partiellement avec les nôtres, il est rare qu'ils coïncident exactement. Dans de nombreuses situations, il est avantageux pour autrui de nous donner de fausses informations ou de nous tromper. C'est parce que ce risque de tromperie existe que la vigilance épistémique peut être non simplement avantageuse, mais même indispensable pour que la communication elle-même reste avantageuse.

Dans la plupart des cas, la communication humaine est intentionnelle et ouverte (*overt*, c'est-à-dire «à visage découvert») le locuteur produit une action par laquelle il communique non seulement une information, mais aussi le fait que, ce faisant, il agit de façon intentionnelle (Grice 1975, Sperber et Wilson 1995).<sup>2</sup> Pour qu'un type de communication ainsi caractérisé puisse être couronné de succès, le locuteur et son destinataire doivent coopérer au prix d'un certain effort. Dans le cas du locuteur, il s'agit de l'effort requis par la production d'une action communicative; dans le cas du destinataire, de l'effort requis pour y prêter attention et l'interpréter. Aucun des deux n'est enclin à consentir cet effort sans attendre, en retour, quelque bénéfice. Pour le destinataire, le bénéfice normalement attendu est l'acquisition d'informations pertinentes et vraies. Pour le locuteur, ce bénéfice est de produire un effet désiré sur le destinataire. Pour satisfaire les attentes du destinataire, le locuteur devrait faire de son mieux pour communiquer une information vraie. Pour satisfaire ses propres attentes, par contraste, il devrait

2 La communication humaine, envisagée en ces termes, est très différente des multiples formes de communication animale discutées par Dawkins et Krebs (1978) et Krebs et Dawkins (1984). En revanche, des considérations évolutionnistes similaires, en termes de coûts et de bénéfices pour les émetteurs et les récepteurs, sont pertinentes dans les deux cas.



choisir de communiquer l'information la plus à même de produire l'effet escompté sur le destinataire, indépendamment de sa valeur de vérité.<sup>3</sup>

Il y a des situations dans lesquelles les émetteurs n'ont aucun intérêt à tromper leur auditoire, par exemple lorsqu'ils enseignent des choses à leurs propres enfants ou lorsqu'ils coordonnent une action collective. Les humains cependant communiquent dans une variété beaucoup plus importante de situations, avec des interlocuteurs dont les intérêts divergent souvent des leurs.

Les bénéfices potentiels de la communication sont immenses, mais ils vont de pair avec le risque d'être induit en erreur, de façon accidentelle ou intentionnelle, ce qui pourrait réduire, annuler ou même inverser ces bénéfices. Le fait que la communication soit aussi omniprésente en dépit de ce risque suggère que les humains sont en mesure de calibrer leur confiance de façon à rendre la communication avantageuse en moyenne à la fois pour les locuteurs et pour les destinataires (Sperber 2001, Bergstrom *et al.* 2006). Pour que cela soit possible, les capacités en œuvre dans la communication explicite et intentionnelle d'une part, dans la vigilance épistémique d'autre part ont dû évoluer en même temps; elles doivent également se développer dans l'individu et être mises à contribution ensemble. Il est donc probable qu'une disposition à la vigilance ait évolué biologiquement parallèlement à la communication telle que les humains la pratiquent. La vie sociale humaine (avec une certaine variabilité culturelle) fournit suffisamment d'inputs pertinents pour le développement de mécanismes psychologiques de vigilance épistémique. Il est vraisemblable en outre que l'interaction entre agents vigilants engendre, outre des mécanismes psychologiques, des mécanismes de vigilance épistémique sociaux. Avant de passer à l'examen d'un certain nombre de ces mécanismes, nous allons considérer quelques questions philosophiques pertinentes pour l'étude de la vigilance épistémique.

3 Il convient de remarquer cependant que si des effets tout autant souhaitables (y compris des effets concernant la réputation sur le long terme) peuvent être obtenus aussi bien par la communication d'informations véridiques que fausses, il pourrait s'avérer préférable de communiquer de l'information véridique, puisque celle-ci est généralement plus facile à gérer (voir Paglieri et Woods 2011, qui développent cet argument de la parcimonie et qui en tirent des conclusions plus fortes que les nôtres).

## 2. Confiance épistémique et vigilance

La confiance est, bien sûr, un aspect essentiel de l'interaction humaine (elle constitue également un vieux sujet philosophique – voir Origgi 2005, 2008). Moins évidente est la thèse selon laquelle les humains sont non seulement amenés à se faire confiance la plupart du temps, mais sont d'emblée confiants et disposés à se croire les uns et les autres, cette confiance fondamentale n'étant suspendue que lorsqu'ils auraient des raisons particulières de se méfier. Cependant, certains philosophes, de Thomas Reid à Tyler Burge et Ruth Millikan, et plus récemment certains psychologues comme Daniel Gilbert ont défendu une version forte de cette thèse en soutenant que les humains sont fondamentalement confiants, pour ne pas dire crédules. Dans cette section, nous discutons ces thèses philosophiques et psychologiques et suggérons que, pour que la confiance puisse jouer le rôle fondamental qui est le sien, elle doit s'appuyer sur une vigilance épistémique active.

Dans l'épistémologie classique, l'acceptation non critique des dires d'autrui était perçue comme incompatible avec ce que la rationalité requiert d'un savoir authentique. Une telle acceptation ne présente en effet pas les mêmes garanties que celles qui accompagnent les idées claires et distinctes ou les impressions sensorielles auxquelles on arrive par soi-même. Selon John Locke, par exemple, «Le flottement des opinions d'autres hommes dans nos cerveaux ne nous rend pas le moins du monde plus informés, même si elles s'avèrent être vraies» (Locke 1960, Livre I, ch. 3, sect. 23). Historiquement, cette position individualiste procède d'une réaction contre le rôle écrasant, dans la scolastique, des arguments d'autorité. Elle est toujours présente en épistémologie contemporaine, où, selon une conception répandue, décrite par Tony Coady (1992) comme 'réductiviste' et par Elisabeth Fricker (1995) comme 'réductionniste', les croyances vraies acquises par le biais du témoignage d'autrui ne peuvent être considérées comme des connaissances que si l'acceptation du témoignage lui-même est justifiée par d'autres croyances vraies ayant été acquises non par le témoignage, mais par la perception ou l'inférence (voir Fricker 1995, Adler 2002, van Cleve 2006).

Cette vision réductionniste contraste avec l'approche 'anti-réductionniste', qui considère que la confiance dans le témoignage est intrinsèquement justifiée (Hardwig 1985, Coady 1992, Foley 1994). Selon

Thomas Reid, qui a articulé le premier une vision anti-réductionniste encore influente, non seulement les humains font confiance aux dires d'autrui, mais ils sont, de surcroît, fondés à le faire. Dieu leur a octroyé une disposition à dire le vrai et une disposition à accepter ce que d'autres personnes leur disent comme vrai. Reid met en avant ces deux principes qui 'vont de pair', le Principe de Vérité et le Principe de Crédulité (Reid 1764, § 24).

Les défenseurs modernes d'une épistémologie reidienne tirent argument de l'existence même du langage naturel pour prouver que des principes de crédulité et de vérité sont effectivement à l'œuvre. Comment pourrait-on expliquer que des significations partagées dans une langue publique aient pu se stabiliser si ce n'est par le fait que la plupart des énoncés dans cette langue sont des témoignages véridiques? Selon Lewis (1969) et Davidson (1984) en particulier, la simple possibilité d'une langue commune présuppose l'emploi d'une parole généralement véridique. Cet argument peut servir à justifier a priori la confiance accordée au témoignage (voir Coady 1992). Ainsi, Tyler Burge soutient le 'caractère purement conservateur' de la communication linguistique: de même que la mémoire est un moyen de conservation de contenu chez l'individu, le langage est un moyen de conservation de contenu d'un individu à l'autre. Dans cette perspective, tout acte de communication implique un engagement tacite envers un principe d'acceptation qui nous autorise à «accepter comme vrai ce qui est présenté comme tel et qui [nous] est intelligible, sauf s'il existe de fortes raisons de ne pas le faire» (Burge 1993, pp. 457-88). Approchant la question sous un angle évolutionniste, Ruth Millikan (1987) soutient que le témoignage est une forme de perception par procuration et que, en tant que tel, il constitue une source directe de connaissance qui ne nécessite pas davantage de justification inférentielle que la connaissance par le moyen de la perception.<sup>4</sup>

Ce débat entre réductionnisme et anti-réductionnisme tourne autour de deux questions distinctes, l'une normative et l'autre descriptive. La question normative porte sur les conditions dans lesquelles une croyance acquise par

4 Il est à remarquer que cette perspective anti-réductionniste, qui traite le témoignage comme un simple processus de transfert de contenu et qui considère les interprètes comme de simples récepteurs de contenu, ignore largement ou nie la nature systématiquement contextuelle et constructive de la compréhension, qui a pourtant bien été établie par la pragmatique moderne. Bezuidenhout (1998) s'oppose à Burge en ces termes, et Origgi et Sperber (2000) critiquent Millikan sur ces mêmes bases.

témoignage est assimilable à une connaissance. La question descriptive concerne les pratiques cognitives et sociales intervenant dans la production et l'acceptation des témoignages. Les deux questions sont explicitement reliées par les défenseurs d'une «troisième voie» selon laquelle nos pratiques effectives, qui supposent un certain degré de vigilance, sont probablement raisonnables et donc nous informent aussi sur la norme (par ex. Adler 2002, Fricker 2006).

La question descriptive a récemment été traitée par la psychologie expérimentale. Les travaux de Daniel Gilbert et ses collègues, en particulier, semblent montrer que, dans un premier temps, nos systèmes mentaux acceptent automatiquement l'information communiquée, avant éventuellement de l'examiner et de la rejeter (Gilbert *et al.* 1990, Gilbert *et al.* 1993). Ces recherches pourraient (d'un point de vue descriptif plutôt que normatif) faire pencher la balance du côté d'une approche anti-réductionniste de la connaissance acquise par témoignage. Dans une expérience représentative de l'ensemble de ces recherches, on demandait aux participants d'apprendre des mots soi-disant «en Hopi». On leur présentait des phrases telles que: «une Monishna est une étoile», immédiatement suivies d'un signal VRAI ou FAUX indiquant la valeur de vérité de la phrase. Dans certains cas, cependant, on déconcentrait les participants pendant qu'ils traitaient le signal VRAI ou FAUX. Un peu plus tard, ils passaient une tâche de reconnaissance dans laquelle les mêmes énoncés sur le sens de mots prétendument en Hopi étaient présentés, et ils avaient à juger si ces énoncés étaient vrais ou faux. La prédiction des auteurs de cette étude était la suivante: si l'acceptation est automatique, le fait d'avoir distrait les participants pendant l'affichage de signaux indiquant que l'énoncé précédent était faux devrait les mener à se remémorer ces énoncés comme étant vrais. Les résultats ont confirmé cette prédiction.

Quelle est la portée des données de Gilbert *et al.*? La vigilance épistémique suppose un coût de traitement qui est probablement maintenu à un niveau minimum lorsque l'information communiquée est sans aucune pertinence pour l'individu. Si, par exemple, vous entendez une information à la radio concernant une compétition sportive dont vous ignorez tout et qui ne vous intéresse pas, vous n'allez probablement pas faire d'effort particulier pour décider si croire ou non ce que vous venez d'entendre. Si on vous met en demeure de dire si l'information était vraie ou fausse, vous direz sans doute

qu'elle était vraie. Après tout, l'information n'a pas été seulement énoncée, elle a aussi été affirmée. Dire que l'information était fautive reviendrait à remettre en question la légitimité de l'affirmation; à quoi bon dans de telles circonstances?

Des expériences plus récentes ont souligné le rôle crucial de la pertinence – ou, plutôt, de la non-pertinence – des données utilisées par Gilbert et ses collègues. Quand bien même les participants pourraient faire preuve d'un quelconque intérêt pour des énoncés à propos du sens de mots Hopi (ce qui n'est guère probable vu la situation expérimentale et les compétences des participants), l'information selon laquelle l'une de ces affirmations (par exemple, 'Une Monishna est une étoile') est fautive serait totalement non pertinente pour eux. Savoir qu'un énoncé de ce type est faux n'entraîne aucune conclusion. Avec d'autres types d'énoncés, les choses peuvent être différentes. Si vous aviez des raisons préalables de penser qu'une affirmation donnée est vraie, ou si celle-ci décrivait un état de choses normal, on voit bien comment il pourrait être pertinent de découvrir qu'elle est fautive. Par exemple, on voit bien comment le fait d'apprendre qu'une affirmation comme «Patrick est un bon père» est fautive pourrait avoir à vos yeux tout un ensemble d'implications plus ou moins convenues. Et effectivement, dans les expériences de Hasson, Simmons et Todorov (2005), par ailleurs similaires à celles de Gilbert, lorsque les participants avaient affaire à des affirmations dont la fausseté revêtait une pertinence potentielle, l'acceptation automatique disparaissait. Ces résultats jettent un doute sur la valeur des données expérimentales invoquées pour montrer que l'information communiquée est automatiquement acceptée comme vraie (voir aussi Bergstrom et Boyer (soumis), Richter *et al.* 2009).

Comme relevé précédemment, les philosophes et psychologues défendant l'idée que les humains sont fondamentalement confiants ne nient pas que, lorsque les circonstances semblent le requérir, les gens adoptent une attitude critique face à l'information communiquée et peuvent finir par la rejeter. Les défenseurs de cette approche ne sont pas obligés de nier qu'une telle attitude critique puisse exploiter des mécanismes cognitifs de vigilance épistémique spécifiques. La vigilance (contrairement à la méfiance) n'est pas le contraire de la confiance; elle s'oppose plutôt à la confiance aveugle (voir également Yamagishi 2001). Pourtant, les philosophes et psychologues dont nous avons discuté les affirmations dans cette section assument que, même

si les humains ne font pas confiance aveuglément, ils ferment au moins la plupart du temps les yeux sur la possibilité d'être victimes de désinformation. Dans les termes de Gilbert, les humains sont confiants «par défaut» (Gilbert *et al.* 1990: 601) et sont disposés à examiner l'information communiquée de manière critique uniquement lorsque les circonstances les motivent à le faire. Leur approche laisse sans réponse la question de savoir comment reconnaître de telles circonstances sans être vigilant en premier lieu.

Il convient également de remarquer que l'idée d'une confiance par défaut est inspirée d'une conception de la cognition inspirée de l'Intelligence Artificielle dans sa première forme et conçue en termes d'organigrammes séquentiels dans lesquels un mécanisme est entièrement inactif tant qu'il n'entre pas en fonction, fonction qu'il remplit ensuite complètement et de façon ininterrompue. Selon une autre façon de voir les choses, plusieurs mécanismes peuvent opérer en parallèle ou en compétition. Par exemple, il se pourrait que chaque comportement communicatif active deux processus distincts chez le destinataire: l'un destiné à identifier la pertinence de l'information communiquée en postulant qu'elle est digne de confiance, et l'autre destinée à évaluer si elle est effectivement digne de confiance. L'un ou l'autre de ces processus pourrait être interrompu par manque d'input adéquat, ou parce que l'un inhibe l'autre, ou encore à la suite d'une distraction. De manière générale, reconnaître l'importance prépondérante de la confiance au sein de la communication humaine ne doit pas nécessairement nous amener à nier ou à minimiser l'importance de la vigilance épistémique.

Voici une analogie susceptible de clarifier comment la confiance épistémique peut coexister avec la vigilance épistémique, et même être renforcée par elle. Lorsque nous marchons dans la rue au milieu d'une foule de gens dont beaucoup passent juste à côté de nous, il existe un risque permanent de collision accidentelle, voire intentionnelle. Pourtant, nous faisons confiance aux gens dans la rue et n'avons aucune hésitation à déambuler parmi eux. Nous n'attendons pas pour autant des autres qu'ils fassent attention alors que nous marchons nous-mêmes sans nous soucier de rien. Nous surveillons la trajectoire des autres et gardons l'œil ouvert pour identifier les individus éventuellement distraits ou agressifs, en ajustant automatiquement notre niveau de vigilance à notre environnement. La plupart du temps, ce niveau de vigilance est suffisamment bas pour rester

inconscient et pour ne pas nous distraire, par exemple, du plaisir d'une promenade, mais il augmente lorsque la situation le requiert. Notre confiance mutuelle dans la rue est largement basée sur une vigilance mutuelle. De manière similaire, dans la communication, ce n'est pas tant que nous puissions faire confiance à autrui en général et que, par conséquent, nous n'ayons besoin d'être vigilants que dans des circonstances rares et spéciales. C'est plutôt que nous ne pourrions pas faire preuve de confiance mutuelle si nous n'étions pas, pour commencer, mutuellement vigilants.

### 3. Compréhension et acceptation

La communication humaine se caractérise, entre autres, par le fait que les communicateurs ont deux buts distincts: être compris et parvenir à faire penser ou agir leur auditoire conformément à ce qu'ils auront compris. Les destinataires de leur côté peuvent comprendre un message sans pour autant l'accepter (qu'il y ait ou non un biais ou une tendance en faveur de l'acceptation).

La compréhension et l'acceptation sont-elles des processus distincts pour d'autres espèces animales? Il y a quelques données limitées qui pourraient suggérer que tel est bien le cas. On a montré expérimentalement que les singes vervets et rhésus ne tiennent pas compte d'un signal d'alerte émanant d'un individu ayant produit dans le passé une série de fausses alertes (Cheney et Seyfarth 1990, Gouzoules *et al.* 1996). Si l'on interprète ces résultats de façon anthropomorphique, cela pourrait indiquer que les singes comprennent le message, mais qu'ils ne l'acceptent pas compte tenu de la non-fiabilité de la source. Cependant, une explication plus parcimonieuse consiste à dire que les signaux d'alerte de cet individu non fiable ne sont pas d'abord interprétés puis rejetés par ses congénères, mais qu'ils sont simplement traités comme du bruit, et par conséquent rejetés d'office. Plus généralement, il n'y a ni preuves ni arguments solides en faveur d'une distinction entre compréhension et acceptation dans la communication non humaine. (Dans tous les cas, même s'il s'avérait que d'autres animaux sociaux exercent une forme quelconque de vigilance épistémique, cela enrichirait

notre compréhension de leur esprit plutôt que d'appauvrir la compréhension du nôtre).

Les philosophes du langage et les pragmaticiens qui se situant dans la tradition d'Austin, Grice et Strawson se sont particulièrement intéressés à la différence entre compréhension et acceptation ainsi qu'à la relation entre ces processus. Austin, par exemple, a distingué «l'assurance d'avoir été compris»<sup>5</sup> (c'est-à-dire «provoquer la compréhension de la signification et de la valeur de la locution») d'un éventail d'autres effets cognitifs ou comportementaux ultérieurs, qu'il a décrits comme «perlocutoires» (Austin 1962: 116, [1970: 124 pour la traduction française]).

Quant à Grice (1957), il a pris l'intention du locuteur de produire un certain effet cognitif ou comportemental qui dépasse la simple compréhension comme point de départ de son analyse du «vouloir-dire» du locuteur:

«[S] a voulu dire quelque chose par x» est (à peu près) équivalent à «[S] a eu l'intention que l'énonciation de x produise quelque effet chez l'auditoire au moyen de la reconnaissance de cette intention». (Grice 1957/1989, p. 220, traduction libre)

Cette analyse, qui a subi de nombreuses révisions et reformulations (par ex. Strawson 1964, Grice 1969/1989, Searle 1969, Schiffer 1972), traite le vouloir dire du locuteur comme un état mental complexe, composé de plusieurs niveaux d'intentions, dont la plus profondément enchâssée est l'intention de faire en sorte que le destinataire pense ou agisse d'une certaine façon. Au-delà de cette intention de base se trouvent deux intentions d'ordre supérieur: que le destinataire reconnaisse l'intention de base, et que la reconnaissance de celle-ci par le destinataire soit au moins en partie sa raison de la satisfaire. En reconnaissant l'intention de base (et donc en satisfaisant l'intention d'ordre supérieur du locuteur de faire en sorte que son intention de base soit reconnue), le destinataire aura compris l'énoncé, qu'il satisfasse ou non par la suite cette intention de base en produisant la réponse souhaitée.

En suivant cette suggestion concernant la relation entre compréhension et acceptation, Sperber et Wilson (1995) construisent leur modèle inférentiel

5 Note du traducteur: nous renvoyons ici à la traduction française de la formulation anglaise, 'the securing of uptake' (Austin, J.L. 1970, Quand dire, c'est faire, trad. par Gilles Lane, éditions du Seuil). Le traducteur de l'ouvrage d'Austin, dans la note 40, p. 177, précise qu'une traduction «plus littérale de *securing of uptake* serait la suivante: s'assurer de ce qu'un autre 'relève' (cela même que j'ai 'jeté dans' mes paroles).»



de la communication autour de l'idée que les locuteurs ont à la fois une intention informative et une intention communicative. Dans leur cadre, le locuteur produit un énoncé (ou d'autres stimuli ostensifs), en ayant par là-même:

- L'intention informative: l'intention d'informer un auditoire de quelque chose.
- L'intention communicative: l'intention d'informer un auditoire de son intention informative.

Ici, l'intention informative correspond à l'intention basique de premier ordre, chez Grice, de susciter une certaine réponse chez l'auditoire, et l'intention communicative correspond, toujours chez Grice, à l'intention d'ordre supérieur de faire en sorte que cette intention basique soit reconnue. Il convient de remarquer que l'intention communicative est elle-même une intention informative de deuxième ordre, qui est satisfaite une fois que l'intention informative de premier ordre est reconnue. Si le destinataire accepte l'autorité (épistémique ou pratique) du locuteur, la reconnaissance de l'intention informative engendrera sa satisfaction, et par conséquent la production de la réponse cognitive ou comportementale souhaitée. Cependant, l'intention communicative peut être satisfaite sans que l'intention informative correspondante ne le soit: en d'autres termes, un auditoire est en mesure de comprendre correctement un énoncé sans accepter ni se conformer à ce qu'il a compris.

L'analyse de la communication faite par Sperber et Wilson diffère de l'analyse gricéenne du vouloir-dire sur deux points. Sperber et Wilson considèrent que seules deux intentions informatives hiérarchiquement reliées sont en jeu, l'intention communicative étant une intention d'ordre supérieur d'informer l'auditoire de la présence d'une intention informative d'ordre inférieur. Ils rejettent ainsi l'idée que le locuteur doit avoir une intention de troisième ordre selon laquelle la reconnaissance, par le destinataire, de l'intention informative doit constituer, au moins en partie, sa raison de la satisfaire.

Pour Grice, cette intention de troisième niveau est essentielle pour distinguer le fait de 'vouloir-dire' quelque chose du fait de 'montrer' quelque chose. Si je vous montre que j'ai un coquillage dans ma poche, votre raison de croire que j'ai un coquillage dans ma poche est le fait de l'avoir vu. Si je vous dis «j'ai un coquillage dans ma poche», votre raison de croire que j'ai un

coquillage dans ma poche est le fait que je vous l'ai dit. En quoi le fait de vous dire quelque chose vous donne-t-il une raison de le croire? Par le simple acte de produire une assertion, le locuteur indique qu'il s'engage à fournir de l'information authentique au destinataire, et il souhaite que la reconnaissance de cet engagement fournisse à celui-ci un motif d'accepter un contenu qu'il n'aurait par ailleurs pas de raisons (suffisantes) d'accepter. En d'autres termes, la production d'une assertion implique typiquement la revendication d'une autorité épistémique suffisante pour espérer une confiance épistémique de la part du destinataire. De manière similaire, la formulation d'une requête implique typiquement la revendication d'une autorité pratique ou morale suffisante pour attendre du destinataire une réponse appropriée à la requête.

Pour autant, la reconnaissance, par l'auditoire, du fait que le locuteur voudrait que son énoncé suscite confiance ou conformité est-elle une propriété intrinsèque de la communication du sens? Grice lui-même a attiré l'attention sur un contre-exemple pertinent, dont il pensait qu'il présentait un problème pour son analyse, bien qu'il ne l'ait mentionné qu'en passant, sans fournir de solution. Lorsqu'un locuteur produit un argument logique, il a typiquement l'intention que son auditoire accepte la conclusion de son argument non pas en vertu de son autorité en tant que locuteur, mais parce que la conclusion découle de ses prémisses:

Conclusion de l'argument:  $p, q$ , par conséquent  $r$  (sur la base de prémisses énoncées au préalable): Si  $E$  (l'énonciateur) a l'intention de faire en sorte que  $D$  (le destinataire) pense que  $r$ ,  $E$  n'attend en revanche pas (ni n'a l'intention) que  $D$  forme la croyance que  $r$  sur la base de l'intention de  $E$  que  $D$  la forme. Les prémisses, et non la confiance en  $E$ , sont censées faire le travail (Grice 1969/1989: 107, traduction libre).

Malgré l'existence de tels contre-exemples, Grice considérait que ses raisons de conserver cette intention de troisième niveau dans son analyse du 'vouloir-dire' étaient déterminantes. Sperber et Wilson, de leur côté, n'analysaient pas le 'sens' mais la 'communication', et ils considéraient qu'il y a un continuum entre 'vouloir dire que' et 'montrer que' qui rend improductive la recherche d'une démarcation stricte. En formulant un argument explicite, par exemple, un locuteur veut dire et montre à la fois que sa conclusion découle de ses prémisses. Même si la description que Grice fait de cet exemple est peu concluante, elle est pertinente pour l'étude de la vigilance épistémique. Elle souligne le contraste entre les cas où un locuteur a l'intention que son destinataire accepte ce qu'il dit en vertu du fait qu'il le dit, et ceux où il attend

du destinataire qu'il accepte ce qu'il dit parce y reconnaît un argument valide. Nous creuserons dans un instant davantage cette distinction entre la vigilance qui porte sur la source de l'information communiquée et celle qui porte sur son contenu.

Clairement, la compréhension du contenu communiqué par un énoncé est une précondition de son acceptation. Cependant, il ne s'ensuit pas que les deux processus se déroulent séquentiellement. En effet, on admet généralement que des considérations liées à l'acceptabilité jouent un rôle crucial dans le processus de compréhension lui-même. Nous pensons que c'est effectivement le cas, mais pas de la façon communément envisagée. Comme nous l'avons mentionné précédemment, de nombreux philosophes ont défendu l'idée que, pour que la communication soit possible, la plupart des énoncés doivent être considérés comme vrais lorsqu'ils sont bien interprétés. Selon Davidson, par exemple, nous devons interpréter un énoncé «d'une manière qui optimise l'accord» et qui révèle «un ensemble de croyances en grande partie consistantes et vraies selon nos propres normes» (Davidson 1984: 137). Une telle «charité interprétative» implique une réticence à réviser ses propres croyances à la lumière de ce que disent les autres. Il s'agit là d'une attitude générale de confiance a priori consistant à tenir pour acquis que ce que veut dire autrui est vrai – mais cette confiance est bien pingre dans la mesure où c'est en dernier lieu à l'interprète de décider de ce qui est vrai (voir cependant Davidson 1986 pour une position plus nuancée).

Il y a une différence entre croire un locuteur parce que vous interprétez ses dires de manière à rendre son énoncé aussi crédible que possible à vos yeux, et croire ce que vous comprenez de son énoncé – même si c'est incompatible avec vos propres croyances, que vous devrez peut-être réviser par la suite – parce que vous lui faites d'emblée confiance. Dans ce dernier cas, l'interprétation n'est pas guidée par une présomption de vérité; par quoi est-elle donc guidée? Selon Sperber et Wilson (1995), elle est guidée par une attente de pertinence.

Même lorsqu'un énoncé est formulé dans votre propre langue, le décodage de sa signification linguistique reste bien en deçà de ce qui est nécessaire pour déterminer son interprétation. Le processus de compréhension combine cette signification linguistique avec de l'information contextuelle et vise une interprétation cohérente avec l'attente de pertinence suscitée par tout énoncé. Selon la théorie de la pertinence – que nous

simplifions ici – chaque énoncé communique une présomption de sa propre pertinence, une pertinence suffisante pour mériter l’attention du destinataire. Le destinataire n’est pas obligé d’accepter cette présomption: après tout, le locuteur pourrait ignorer ce qui est pertinent pour lui, ou il pourrait tout simplement n’y attacher aucune importance. Mais, que le destinataire accepte ou non la présomption de pertinence, le simple fait que celle-ci soit communiquée est suffisant pour guider le processus d’interprétation. Ce fait justifie la recherche d’une interprétation que le locuteur avait des raisons de penser pertinente pour le destinataire. Dans de nombreux cas, l’output d’une telle interprétation, centrée sur des considérations de pertinence, diffère de celui que la charité interprétative favoriserait.

Pour illustrer ce point, supposons que Barbara a demandé à Julie d’amener une bouteille de champagne à la fête:

- (1) André (à Barbara): Une bouteille de champagne? Mais le champagne, c’est cher!
- (2) Barbara: Julie a de l’argent.

André avait précédemment tenu pour acquis que Julie était une simple universitaire mal payée en début de carrière. Comment devrait-il alors interpréter la réponse de Barbara? Si son but était d’optimiser l’accord entre ce qu’il croit et ce que dit Barbara, il devrait considérer que Barbara veut dire que Julie a de l’argent, par opposition à ne pas avoir d’argent du tout, ce qui est vrai de la plupart des gens et chose qu’il croyait déjà de Julie. Mais une telle interprétation ne rend pas la réponse de Barbara assez pertinente pour mériter l’attention d’André. Par contraste, s’il interprète «a de l’argent» comme voulant dire «a assez d’argent pour être en mesure d’acheter du champagne», l’énoncé de Barbara sera suffisamment pertinent. Plus précisément, l’énoncé sera suffisamment pertinent pour André à condition qu’il le croie. S’il ne croit pas ce qu’il a compris de l’énoncé de Barbara, alors cet énoncé ne lui fournira rien d’autre que des informations sur Barbara elle-même (ses croyances, son vouloir-dire), plutôt que des informations sur Julie, ce qui pourrait bien ne pas s’avérer assez pertinent pour lui. Comment, dès lors, des considérations de pertinence aident-elles André à comprendre le sens de l’énoncé de Barbara alors même qu’il n’est pas disposé à l’accepter? Notre thèse est la suivante: qu’il finisse par l’accepter ou non, le destinataire d’un énoncé attribue à l’émetteur l’intention d’asserter une proposition qui serait assez pertinente pour lui s’il l’acceptait.

En d'autres termes, les destinataires doivent adopter une 'posture de confiance' au cours de leur interprétation (voir Holton 1994; Origg 2005, 2008). La confiance requise est moins 'pingre' que celle requise par le principe de charité. Elle implique chez le destinataire une disposition à ajuster ses propres croyances à une interprétation du vouloir-dire du locuteur guidée par la pertinence, par opposition à un ajustement de sa propre interprétation du vouloir-dire du locuteur à ses propres croyances. D'un autre côté, il s'agit d'une confiance provisoire. Interpréter un énoncé comme si on allait l'accepter n'est pas équivalent à l'accepter dans les faits, pas même à l'accepter 'par défaut'.

Cela dit, il se pourrait que la posture de confiance que la compréhension implique cause, ou contribue à causer, une tendance ou un biais en faveur de l'acceptation réelle de l'information communiquée. Si tel est le cas, cela pourrait contribuer à expliquer les résultats des expériences psychologiques de Gilbert et les considérations introspectives qui conduisent les philosophes reidiens à supposer (à tort) que la confiance épistémique est une disposition par défaut. Remarquons qu'une simple tendance ou biais en faveur de l'acceptation de l'information communiquée ne serait pas irrationnelle étant donné que, en fait, la communication est la plupart du temps honnête (et nous maintenons que c'est le cas notamment parce que la vigilance de l'auditoire limite l'éventail des situations où la malhonnêteté serait dans le meilleur intérêt des locuteurs).

Somme toute, comprendre n'est pas équivalent à croire mais n'équivaut pas non plus à adopter une position sceptique. La compréhension implique l'adoption d'une attitude de confiance provisoire et fragile; cette posture mènera à l'acceptation de ce qui est compris seulement si la vigilance épistémique, qui est déclenchée par les mêmes actes communicationnels qui déclenchent la compréhension, ne fournit pas de raisons de douter.

#### 4. Vigilance envers la source

La communication apporte des bénéfices vitaux mais comporte pour l'auditoire le risque majeur d'être accidentellement ou intentionnellement

désinformé. Il n'existe aucune méthode infaillible pour calibrer sa propre confiance dans l'information communiquée de manière à évacuer toutes les informations erronées et uniquement celles-là. Comme les enjeux sont particulièrement élevés, il est plausible de penser qu'une pression sélective continue s'est exercée en faveur de tout moyen disponible et, même approximatif, permettant un tel tri à un coût acceptable. Compte tenu de l'existence d'une variété de considérations pertinentes pour l'octroi ou la rétention de la confiance épistémique, nous explorerons la possibilité que différentes compétences favorisant la vigilance épistémique aient émergé au cours de l'évolution biologique et culturelle, chacune étant spécialisée dans un type particulier de considérations pertinentes.

Les facteurs qui influencent l'acceptation ou le rejet d'une information communiquée peuvent concerner soit la source de l'information – qui croire?, soit son contenu – que croire? Nous allons dans un premier temps considérer la vigilance épistémique qui porte sur la source de l'information.

Les jugements que l'on porte sur la fiabilité d'un informateur peuvent être plus ou moins généraux ou contextualisés. On peut par exemple penser de Marie «c'est une personne de confiance», l'entendant à la fois épistémiquement et moralement, et par conséquent s'attendre à ce que Marie dise la vérité, à ce que ce qu'elle accomplisse soit moralement bon, et ainsi de suite. On peut faire confiance à quelqu'un (ou s'en méfier) sur un sujet particulier dans des circonstances spécifiques: par exemple, on peut de manière générale faire confiance à Julie pour tout ce qui touche aux estampes japonaises, mais rester vigilant lorsqu'elle en vend une elle-même. La confiance peut certes être allouée de ces deux manières; mais en quoi celles-ci diffèrent-elles d'un point de vue normatif?

Un informateur fiable doit satisfaire deux conditions: il doit être compétent et il doit être bienveillant. C'est-à-dire qu'il doit posséder de l'information authentique (par opposition à de l'information erronée ou à un manque total d'information) et qu'il doit avoir l'intention de partager cette information authentique avec son auditoire (par opposition à produire des assertions qu'il considère lui-même comme fausses, que ce soit par indifférence ou malveillance). Evidemment, le même informateur peut être compétent sur un sujet mais pas sur d'autres, et bienveillant envers un certain auditoire dans certaines circonstances, mais pas envers un autre auditoire dans d'autres circonstances. Cela suggère que la confiance devrait être

accordée aux informateurs en fonction des sujets, des auditoires et des circonstances. Cependant, une telle précision dans le calibrage de la confiance est une opération coûteuse en termes cognitifs et, s'il arrive souvent que les gens soient disposés à payer ce coût, ils s'appuient plus communément sur des impressions générales de compétence, de bienveillance et de fiabilité globale beaucoup moins coûteuses.

Une illustration frappante de la tendance que nous avons à former des jugements généraux de confiance sur la base d'indices très limités est fournie par une étude de Willis et Todorov (2006). Dans cette expérience, on montrait aux participants des photographies de visages, soit pendant 100 millisecondes seulement, soit sans limite de temps, puis on leur demandait d'évaluer la fiabilité, la compétence, le caractère sympathique, l'agressivité et l'attrait des personnes dont ils avaient vu les visages. Contrairement aux attentes des auteurs, la corrélation entre jugements avec et sans limite de temps n'était pas plus élevée pour l'attrait (.69) – qui est, après tout, une propriété de l'apparence d'une personne – que pour la confiance (.73), alors que les corrélations pour l'agressivité et la compétence restaient relativement basses (.52). Il est permis de douter que de tels jugements de confiance, formés en une fraction de seconde, soient basés sur quoi que ce soit de valable, mais ce que cette expérience suggère fortement, c'est que le fait de rechercher des signes de fiabilité est l'une des premières choses que l'on fait lorsqu'on voit un nouveau visage (voir aussi Ybarra *et al.* 2001).

Beaucoup de travaux en psychologie sociale suggèrent que le comportement des gens est, pour une part significative, déterminé non pas par leur personnalité, mais par la situation (Ross & Nisbett 1991, Gilbert & Malone 1995). Si tel est le cas, le fait de juger que quelqu'un est généralement digne de confiance pourrait relever d'une telle «erreur fondamentale d'attribution» (Ross 1977): à savoir, de la tendance à surestimer le rôle des dispositions psychologiques et à sous-estimer les facteurs situationnels lorsqu'il s'agit d'expliquer ou de prédire le comportement de quelqu'un. Mais même sans avoir recours à une psychologie des traits de caractère, il est possible de défendre l'idée que certaines personnes sont généralement plus dignes de confiance que d'autres, et qu'elles sont dans une certaine mesure reconnaissables comme telles.

Si nous interagissons continuellement avec les mêmes personnes, mal les informer lorsque cela sert notre intérêt immédiat est potentiellement

dommageable pour notre réputation, et par conséquent coûteux sur le long terme. À l'inverse, faire de notre mieux pour être systématiquement digne de confiance peut s'avérer parfois coûteux sur le court terme, mais bénéfique sur le long terme. La pondération des coûts à court terme et des bénéfices à long terme dans le choix de se comporter de façon à mériter la confiance d'autrui peut différer d'une personne à l'autre selon, par exemple, l'importance qu'elles accordent au futur plus ou moins éloigné (Ainslie 2001), ce qui peut entraîner des choix différents en la matière. Si les humains diffèrent en effet dans le degré auquel ils sont disposés à payer le coût de leur propre fiabilité, il est possible que des jugements généraux sur la crédibilité relative des uns et des autres ne soient pas sans fondement. Les gens qui optent pour une politique de fiabilité systématique récolteraient ainsi les bénéfices d'une réputation de personne hautement fiable. Une telle réputation, nourrie par une connaissance partagée de leurs actions passées, peut de surcroît être soulignée au quotidien par leur comportement et leur attitude.

Il est donc possible de projeter une image de crédibilité (que cette image soit elle-même crédible ou pas). Est-il également possible, à l'inverse, d'agir de façon à tromper autrui, en mentant par exemple, sans n'en laisser rien paraître? La littérature sur la détection du mensonge est substantielle (voir Ekman 2001, pour un panorama); elle montre, en deux mots, que le fait de détecter les mensonges sur la base d'indices comportementaux non-verbaux est une chose difficile (Vrij 2000, Malone & DePaulo 2001, Bond & DePaulo 2006), même pour des gens entraînés à la tâche (par ex. DePaulo & Pfeifer 1986, Ekman & O'Sullivan 1991, Mann, Vrij & Bull 2004, Vrij 2004), et même lorsque les menteurs sont loin d'être des experts – par exemple, lorsque ce sont des enfants de trois ans (Lewis *et al.* 1989, Talwar & Lee 2002). L'aptitude à mentir peut s'avérer fort avantageuse, mais uniquement si le menteur ne se trahit pas lui-même. Quelles que soient les contributions respectives des dispositions héritées de l'évolution et des compétences acquises, les menteurs semblent à même de maintenir les signes comportementaux de malhonnêteté à un strict minimum.

Afin de mieux comprendre les mécanismes de vigilance épistémique qui portent sur la source, nous avons un besoin urgent non pas de plus de recherches empiriques sur la détection du mensonge ni sur les jugements généraux de crédibilité, mais plutôt de davantage d'études sur la façon dont



la confiance et la méfiance sont calibrées par rapport à la situation, aux interlocuteurs et au thème de la communication. Deux types distincts de considérations devraient ici être pris en compte: la compétence de l'émetteur sur le sujet dont il parle d'une part, et sa motivation à communiquer d'autre part. La compétence et l'honnêteté sont toutes deux des conditions de crédibilité. Il existe une littérature considérable, indirectement pertinente pour l'étude de la vigilance épistémique dans la communication ordinaire, par exemple en histoire et sociologie des sciences (par ex. Shapin 1994), en anthropologie du droit (par ex. Hutchins 1980, Rosen 1989), dans l'étude linguistique des expressions évidentielles (par ex. Chafe & Nichols 1986, Ifantidou 2001, Aikhenvald 2004), ou en psychologie sociale de l'influence et de la persuasion (par ex. Chaiken 1980, Petty et Cacioppo 1986). Cependant, de plus amples recherches sur la vigilance épistémique dans la communication quotidienne doivent encore être entreprises.

Dans la section suivante, nous nous tournons vers le développement de la vigilance épistémique au cours de l'enfance, thème qui est non seulement intéressant en lui-même, mais qui nous aidera en outre à séparer les différentes dimensions de la vigilance épistémique vis-à-vis de la source de l'information.

## 5. Le développement de la vigilance épistémique (et de la lecture de l'esprit)<sup>6</sup>

Les études sur le développement de la vigilance épistémique chez l'enfant se multiplient (voir par ex. Koenig & Harris 2007, Heyman 2008, Clément 2010, Corriveau & Harris 2010, Nurmsoo *et al.* 2010). Ces études montrent que même à un âge très précoce, les enfants n'attribuent pas un degré de fiabilité égal à toutes les informations communiquées. A 16 mois, ils remarquent

6 Note du traducteur: nous avons opté pour une traduction littérale du terme anglais, *mindreading*, en suivant la littérature francophone disponible dans le domaine. Par 'lecture de l'esprit' nous entendons donc le processus par lequel on se représente les états mentaux d'autrui.

l'emploi inapproprié d'un mot familier (Koenig & Echols 2003). A 2 ans, ils essaient souvent de contredire et de corriger des affirmations qu'ils croient fausses (par ex. Pea 1982). Ces recherches remettent en cause l'idée répandue selon laquelle les jeunes enfants sont simplement crédules.

Les jeunes enfants disposent-ils de ressources cognitives suffisantes pour accorder leur confiance sur la base d'informations pertinentes quant à la fiabilité d'un informateur? Lorsqu'ils ont le choix, les enfants de 3 ans semblent préférer des informateurs qui sont à la fois bienveillants (Mascaro & Sperber 2009) et compétents (par ex. Clément *et al.* 2004). Lorsqu'ils préfèrent des informateurs bienveillants, ils tiennent compte non seulement de leurs propres observations, mais aussi de ce qu'on leur a dit du caractère moral de l'informateur (Mascaro & Sperber 2009), et lorsqu'ils préfèrent des informateurs compétents, ils tiennent compte de leur exactitude passée (par ex. Clément *et al.* 2004, Birch *et al.* 2008, Scofield & Behrend 2008). A l'âge de 4 ans, non seulement ils ont des préférences appropriées pour les informateurs fiables, mais ils font aussi preuve d'une certaine compréhension de ce que la fiabilité implique. Par exemple, ils peuvent prédire qu'un informateur malhonnête va fournir une information fautive (Couillard & Woodward 1999), ou qu'un informateur incompetent sera moins fiable (Call & Tomasello 1999, Lampinen & Smith 1995, Clément *et al.* 2004). De plus, ils font de telles prédictions malgré le fait que les informateurs non fiables se présentent typiquement comme bienveillants et compétents.

La vigilance épistémique précoce exploite certaines compétences utilisées dans la sélection de partenaires pour la coopération, parmi lesquelles l'évaluation morale, le suivi de la fiabilité et la vigilance envers les tricheurs (par ex. Cosmides & Tooby 2005, Harris & Núñez 1996). L'exercice de la vigilance épistémique non seulement s'appuie sur certains de ces processus qui sous-tendent la sélection de partenaires coopératifs, mais contribue aussi à leur efficacité. En particulier, cet exercice contribue à la fiabilité relative des systèmes de réputation, un outil fondamental pour la sélection de partenaires coopératifs (par ex. Alexander 1987, Nowak & Sigmund 1998, Milinski *et al.* 2002 – mais voir ci-après). Par exemple, à mesure que les enfants de 4 et 5 ans deviennent plus vigilants envers la tromperie (Couillard & Woodward 1999, Mascaro & Sperber 2009), ils deviennent également plus vigilants envers l'hypocrisie dans la présentation de soi (Peskin 1996, Mills & Keil 2005, Gee & Heyman 2007).

La vigilance épistémique dirigée vers les informateurs génère une variété d'attitudes épistémiques (acceptation, doute ou rejet, par exemple) envers les contenus communiqués par ces informateurs. Des expériences tendent à montrer que les enfants de 3 ans reconnaissent des attitudes comme l'approbation ou le doute (Fusaro & Harris 2008), et qu'ils ont également conscience du fait qu'une affirmation peut être plus ou moins forte (Sabbagh & Baldwin 2001, Birch *et al.* 2008, Matsui *et al.* 2009). Les enfants sont capables d'interpréter des commentaires sur la fiabilité de ce qui est communiqué (par ex. Fusaro & Harris 2008, Clément *et al.* 2004). Ils peuvent par ce moyen profiter des jugements épistémiques d'autrui et enrichir leur propre compréhension épistémique ainsi que leur capacité de vigilance épistémique.

Les enfants semblent aussi disposer d'une certaine capacité à comparer la fiabilité de différentes sources d'information. Dans des expériences s'inspirant des études célèbres de Salomon Asch sur le conformisme (Asch 1956), par exemple, on s'aperçoit qu'une majorité d'enfants de 3 ans font confiance à leurs propres perceptions plutôt qu'à une série cohérente de jugements erronés produits par des comparses des expérimentateurs (Corriveau & Harris 2010; mais voir Walker & Andrade 1996). Les enfants tiennent compte de l'accès à l'information qu'ont les informateurs (par ex. Robinson & Whitcombe 2003, Robinson *et al.* 2008, Nurmsoo et Robinson 2009). Ils attribuent aux autres des dispositions durables de plus ou moins grande fiabilité (par ex. Koenig & Harris 2007, Birch *et al.* 2009, Corriveau & Harris 2009) et il s'appuient ce faisant sur leur compréhension du fait que différentes personnes sont plus ou moins bien informées – une composante de la psychologie naïve de l'enfant sur laquelle on n'a pas enquêté en profondeur (mais voir Lutz & Keil 2002). La vigilance épistémique des enfants a donc recours à différents aspects de leur épistémologie naïve (sur lesquels, du coup, elle nous informe): leur compréhension du fait que l'accès à l'information, la force des croyances, le fait d'être bien informé et l'engagement vis-à-vis des assertions peuvent tous être plus ou moins forts.

Lorsque la vigilance épistémique est dirigée vers le risque de tromperie, elle requiert une compréhension non seulement des attitudes épistémiques de l'émetteur, mais également de ses intentions, y compris l'intention d'induire de fausses croyances chez son auditoire. Cela demande une lecture de l'esprit relativement sophistiquée, qui fait usage de métareprésentations

d'ordre supérieur ('Elle croit que non-P mais veut me faire croire que P' combine une attribution de croyance de premier ordre avec une attribution d'intention de deuxième ordre).

Il y a des parallèles intéressants à tirer entre le développement de la vigilance épistémique et les résultats des tâches de fausse croyance traditionnellement utilisés pour mesurer le développement de la lecture de l'esprit. On trouve des rudiments de vigilance épistémique chez les jeunes enfants, probablement même chez les bébés. Cependant, dès l'âge de 4 ans, il y a une transition majeure dans la vigilance épistémique, à la fois envers la malhonnêteté (Mascaro & Sperber 2009; voir également Couillard & Woodward 1999, Jaswal *et al.* 2010) et envers l'incompétence (Povinelli & DeBlois 1992, Call & Tomasello 1999, Welch-Ross 1999, Figueras-Costa & Harris 2001). A 4 ans, les enfants commencent à faire preuve d'une attention plus soutenue envers la qualité épistémique des croyances et des messages d'autrui. Ils deviennent beaucoup plus sélectifs dans leur confiance et se montrent de surcroît beaucoup plus aptes et disposés à manipuler les croyances des autres.

Une telle transition dans la vigilance épistémique se produit autour de l'âge auquel les enfants réussissent des tâches de fausse croyance standard (Wimmer & Perner 1983, Baron-Cohen *et al.* 1985). Jusqu'à récemment, cette convergence aurait pu être interprétée de la manière suivante: lorsque les enfants ont environ 4 ans, à la suite de l'émergence de la compréhension de ce qu'est une croyance – et une fausse croyance en particulier – ils deviennent de plus en plus conscients que d'autres personnes peuvent avoir des croyances erronées. Cette conscience constitue une base pour des formes plus matures de vigilance épistémique. Bien qu'elle soit séduisante, cette ligne d'interprétation est devenue moins plausible à la suite d'expériences récentes sur des versions non-verbales de la tâche de fausse croyance adaptée pour des bébés. Ces expériences suggèrent que dès leur seconde année, les enfants s'attendent à ce que le comportement d'un agent soit guidé par ses croyances, même lorsque celles-ci sont fausses (par ex. Onishi & Baillargeon 2005, Southgate *et al.* 2007, Surian *et al.* 2007). Si tel est le cas, les résultats robustes d'études faites avec la tâche de fausse croyance standard doivent être réinterprétés, de même que la transition, mentionnée précédemment, qui a lieu autour de l'âge de 4 ans.

Il est aisé d'imaginer deux réinterprétations possibles de cette transition (et une perspective complète peut émerger de ces deux interprétations). Premièrement, la capacité de réussir des tâches de fausse croyance standard et la capacité améliorée de vigilance épistémique pourraient avoir une cause commune, par exemple, un développement majeur des fonctions exécutives (par ex. Perner & Lang 2000, Carlson & Moses 2001). Deuxièmement, consécutivement à une amélioration de leur capacité de vigilance épistémique, les enfants pourraient commencer à prêter davantage attention aux aspects pertinents des tâches de fausse croyance qui sont généralement ignorés à un plus jeune âge. À mesure qu'ils deviennent de plus en plus conscients que d'autres personnes peuvent entretenir des croyances erronées (résultant d'une malchance épistémique ou de tromperie), ils deviennent également meilleurs dans la prise en compte de ces croyances erronées lorsqu'ils cherchent à prédire le comportement d'autrui. Dans ce cas, leur point de vue n'est pas tant celui d'un observateur, mais plutôt celui d'une potentielle victime de désinformation, celui d'un auteur potentiel de manipulation, ou celui d'un coopérateur doté d'une préférence pour des partenaires bien informés.

Les recherches actuelles sur la vigilance épistémique offrent par conséquent des éclairages intéressants sur la nature et le développement des capacités liées à la théorie de l'esprit. Elles montrent que la vigilance épistémique convoque une variété de mécanismes cognitifs ayant des trajectoires développementales différentes, y compris le sens moral relatif à la reconnaissance de partenaires potentiels de coopération, l'épistémologie naïve et la lecture de l'esprit.

## 6. Vigilance envers le contenu

Comme nous l'avons vu dans les deux sections précédentes, la vigilance épistémique peut être dirigée vers la source de l'information communiquée: l'émetteur est-il compétent et honnête? La vigilance épistémique peut également être dirigée vers le contenu de la communication, qui peut être plus ou moins crédible indépendamment de sa source. Dans cette section et

dans la suivante, nous considérons la vigilance épistémique dirigée vers le contenu de la communication.

Certains contenus sont intrinsèquement crédibles, même s'ils proviennent d'une source non fiable. Les tautologies, les preuves logiques, les truismes et les contenus dont la vérité est mise en évidence par l'acte de communication lui-même (par exemple, dire *'I am able to say a few words in English'*) en sont des exemples. D'autres contenus sont intrinsèquement non crédibles même s'ils proviennent d'une source fiable. Les contradictions logiques, les faussetés flagrantes, et les contenus dont la fausseté est mise en évidence par l'acte de communication lui-même (par exemple, dire *'Je suis muet'*) en sont des exemples.

Dans la plupart des cas, cependant, la vigilance épistémique dirigée vers le contenu communiqué doit s'appuyer sur davantage que les propriétés logiques inhérentes des énoncés, le savoir indubitable d'arrière-plan ou le fait que l'énoncé soit en mesure de confirmer ou d'infirmer son propre contenu. La crédibilité de l'information nouvellement communiquée doit être évaluée par rapport à des croyances d'arrière-plan qui sont elles-mêmes sujettes à révision. Evidemment, les nouvelles informations ne peuvent pas être évaluées relativement à l'intégralité de notre 'encyclopédie mentale'. Pour maintenir un coût et un temps de traitement dans des limites acceptables, seul un très petit sous-ensemble de cette encyclopédie, étroitement lié à la nouvelle information, peut être amené à jouer un rôle dans son évaluation. En effet, l'activation systématique de ne serait-ce qu'un sous-ensemble restreint d'information d'arrière-plan uniquement dans le but d'évaluer la crédibilité du contenu communiqué serait déjà sensiblement coûteuse en termes de traitement. Nous défendons l'idée qu'une telle activation *ad hoc* n'est pas nécessaire.

D'après la théorie de la pertinence (Sperber & Wilson 1995, 2005, Carston 2002, Wilson & Sperber 2004), le processus de compréhension lui-même suppose l'activation automatique d'informations d'arrière-plan dans un contexte dans lequel l'énoncé peut être interprété comme pertinent. Selon cette hypothèse, les coûts de traitement tendent à être proportionnels au bénéfice cognitif que le système en tire. Nous soutenons que les mêmes informations d'arrière-plan qui sont utilisées dans la recherche de pertinence peuvent également générer une évaluation épistémique certes imparfaite, mais d'un rapport coût/bénéfice adéquat. De plus, comme nous nous

apprêtons à le montrer, la recherche de pertinence implique des étapes inférentielles qui fournissent la base de cette évaluation.

Sperber & Wilson (1995) distinguent trois types d'effets contextuels faisant qu'une nouvelle information peut s'avérer pertinente dans un contexte de croyances préalables. (i) Lorsque la nouvelle information et les croyances contextuelles sont combinées en tant que prémisses, elles peuvent générer des 'implications contextuelles' (implications que l'on ne peut dériver ni du contexte seul ni de la nouvelle information prise isolément) qui seront acceptées comme des nouvelles croyances. (ii) Le degré de confiance de l'individu dans ces croyances contextuellement activées peut croître ou décroître à la lumière de la nouvelle information. (iii) L'information nouvelle peut contredire des croyances activées contextuellement et cela peut déclencher leur révision. Les trois types d'effets contextuels (acceptation de nouvelles croyances contextuellement impliquées, modification de la force des croyances et révision de croyances) tendent à contribuer à l'amélioration des connaissances de l'individu.

Qu'arrive-t-il lorsque le résultat du traitement d'une nouvelle information dans un contexte de croyances existant prend la forme d'une contradiction? Lorsque la nouvelle information a été acquise de manière perceptuelle, il est généralement justifié de faire davantage confiance à ses propres perceptions qu'à sa mémoire et d'adapter ses croyances en conséquence. Par exemple, si vous croyiez que Julie était dans le jardin et que vous l'entendez parler dans le salon, vous mettrez automatiquement à jour votre croyance au sujet de l'endroit où Julie se trouve. Un tel mécanisme de mise à jour automatique est probablement la seule forme de révision de croyance présente chez les espèces animales non-humaines.

Lorsque la nouvelle information a été communiquée, en revanche, trois possibilités sont à envisager. (i) Si la source n'est pas considérée comme une source crédible, la nouvelle information peut simplement être rejetée comme si elle était fausse, et donc non pertinente: tel est le cas, par exemple, lorsqu'un ivrogne dans la rue vous dit qu'il y a un éléphant blanc au coin de la rue. (ii) Si la source jouit d'une certaine autorité et que les croyances d'arrière-plan qui sont en conflit avec ce que la source vous a dit ne sont pas entretenues avec un degré élevé de conviction, celles-ci peuvent être directement corrigées: par exemple, en regardant Gilles, vous aviez pensé qu'il avait tout juste la vingtaine; or il vous dit qu'il a 29 ans. Vous acceptez

l'information comme vraie et pertinente – pertinente en premier lieu parce qu'elle vous permet de corriger vos croyances erronées. (iii) Si vous faites confiance à la source et à vos propres croyances, alors la révision de croyances devient inévitable. Vous devez réviser soit vos croyances d'arrière-plan, soit votre croyance que la source est crédible, même si le choix entre les deux options n'est pas immédiatement transparent. Par exemple, il vous semblait que Gilles avait tout juste la vingtaine; Lucie vous dit qu'il doit être dans la trentaine. Devriez-vous conserver votre propre estimation ou faire confiance à celle de Lucie?

Les choses ne sont pas si différentes lorsque le résultat du traitement de l'information communiquée par une source à laquelle vous faites confiance n'est pas une inconsistance logique, mais une incohérence empirique: c'est-à-dire, lorsque la nouvelle information est incompatible avec certaines de vos croyances d'arrière-plan, étant donnée la présence d'autres croyances d'arrière-plan qui sont davantage enracinées. Par exemple, vous croyiez que Gilles était docteur; Lucie vous informe qu'il n'a que 22 ans. Vous êtes convaincu que, pour devenir docteur, il faut de nombreuses années d'études, et qu'il est par conséquent presque impossible d'être docteur à 22 ans. Vous devriez donc soit refuser de croire Lucie, soit abandonner la croyance que Gilles est un docteur. Afin de préserver la cohérence de vos croyances, vous devez réduire soit votre confiance dans la source, soit votre confiance dans vos croyances les moins fermes.

Le rôle de la vérification de cohérence dans la révision de croyances a été souligné par des philosophes comme Gilbert Harman (1986) et Paul Thagard (2002). Ici, cependant, nous envisageons la vérification de la cohérence non pas comme une procédure épistémique générale de révision de croyances, mais comme un mécanisme de vigilance épistémique dirigé vers le contenu communiqué, mécanisme qui tire parti des informations d'arrière-plan activées en quantité limitée par la procédure de compréhension elle-même.

Ne serait-ce que pour des raisons d'efficacité, on pourrait s'attendre à ce que le type de vérification de cohérence utilisé dans la vigilance épistémique ne requière rien de plus que les révisions minimales nécessaires à ré-établir la cohérence. Dans certains cas, la cohérence est rétablie plus facilement par une méfiance envers la source, et dans d'autres en révisant certaines de nos propres croyances d'arrière-plan. A moins qu'une option ne domine cette compétition au point d'inhiber la conscience d'alternatives, c'est typiquement



une décision consciente qui résout le problème. Le fait de prendre une telle décision implique une pensée d'ordre supérieur, ou métareprésentationnelle, au sujet de nos propres croyances.

Ce que nous suggérons, c'est que la recherche d'une interprétation pertinente, qui est une partie intégrante du processus de compréhension, suppose automatiquement la génération d'inférences susceptibles de faire apparaître des inconsistances ou des incohérences pertinentes pour l'évaluation épistémique. Lorsque de telles inconsistances ou incohérences se produisent, elles déclenchent une procédure entièrement dédiée à une telle évaluation. Il n'en reste pas moins que la compréhension, la recherche de pertinence et l'évaluation épistémique sont des aspects interreliés d'un unique processus global dont le but est de tirer le meilleur parti de l'information communiquée.

## 7. Vigilance épistémique et raisonnement

Considérons maintenant les choses du point de vue de l'émetteur. Supposons qu'il soupçonne que son interlocuteur n'acceptera ce qu'il dit sur la seule base de la confiance et qu'il va probablement exercer une certaine vigilance épistémique en vérifiant dans quelle mesure l'énoncé qu'il traite est cohérent avec ses propres croyances. La vigilance active du destinataire constitue un obstacle pour l'accomplissement du but de l'émetteur. Pourtant, du point de vue de ce dernier, un destinataire vigilant vaut toujours mieux qu'un destinataire qui rejette d'office son témoignage. Et, effectivement, le fait que le destinataire s'appuie sur la cohérence en tant que critère d'acceptation ou de rejet de son énoncé peut offrir au locuteur la possibilité de convaincre son interlocuteur en réussissant à vaincre ses défenses.

Nous avons suggéré que la vérification de la cohérence prend place dans le contexte restreint des croyances utilisées dans la recherche d'une interprétation pertinente de l'énoncé. Mais le destinataire est susceptible d'avoir d'autres croyances moins activées qui auraient pu peser dans la balance en faveur de l'information qu'il est réticent à accepter, pour autant qu'il ait été en mesure d'en tenir compte. Dans ce cas, il peut valoir la peine,

pour le locuteur, de rappeler au destinataire de telles croyances d'arrière-plan, augmentant ainsi l'acceptabilité de son énoncé. Il se peut aussi qu'il y ait d'autres informations que le destinataire accepterait sur la base de la confiance qu'il a dans le locuteur, informations qui seraient cohérentes avec son énoncé et qui le rendraient ainsi plus acceptable.

Pour illustrer notre propos, nous adaptions ci-dessous un exemple célèbre de Grice (1989: 32). André et Barbara sont à Boston et parlent de leur ami Steve:

- (3) André: Steve n'a pas l'air d'avoir de petite amie en ce moment.
- (4) Barbara: Il s'est rendu à de nombreuses reprises à New York dernièrement.

Barbara croit que Steve a une nouvelle petite amie, mais pressent que si elle se contentait de l'affirmer, André, qui vient d'exprimer une opinion contraire, ne serait pas convaincu. Or elle a remarqué que Steve s'est rendu à de nombreuses reprises à New York dernièrement et considère cela comme une indication que Steve a une petite amie dans cette ville. André pourrait aussi s'être rendu compte de ces visites, et même si ce n'est pas le cas, il est susceptible de croire Barbara sur parole. Une fois qu'il aura tenu compte de ces visites, la conclusion que Steve est susceptible d'avoir une petite amie pourrait devenir beaucoup plus acceptable à ses yeux.

Barbara ne cache pas qu'elle souhaite qu'André accepte sa conclusion. Bien au contraire, l'assertion selon laquelle Steve s'est rendu à de nombreuses reprises à New York récemment ne satisfera les attentes de pertinence d'André (ou les attentes de coopération, dans un sens gricéen) que dans la mesure où elle est comprise comme implicite que Steve peut tout à fait avoir une petite amie malgré les doutes d'André. Même si André reconnaît que cette implicature fait partie du vouloir-dire de Barbara, il peut ne pas l'accepter. Ce sur quoi compte Barbara pour le convaincre, ce n'est pas la capacité d'André à comprendre son énoncé, mais sa capacité à saisir la force de l'argument dont les prémisses comprennent l'énoncé littéral et d'autres informations d'arrière-plan (au sujet des raisons probables pour lesquelles Steve s'est régulièrement rendu à New York dernièrement), et dont la conclusion est l'implicature de Barbara.

Dans un scénario légèrement différent, André lui-même remarque que Steve s'est rendu à New York à de nombreuses reprises dernièrement; mais, parce qu'il ne fait pas le rapprochement, il ajoute 'Il n'a pas l'air d'avoir de

petite amie ces temps-ci'. Dans ce cas, Barbara peut souligner ce rapprochement afin de l'aider à arriver à la conclusion visée en disant: 'S'il va à New York, c'est peut-être pour y voir une petite amie', ou 'S'il avait une petite amie à New York, cela expliquerait ses visites'. Elle pourrait encore simplement répéter le commentaire d'André, mais avec un changement dans la tournure: 'Il n'a PAS L'AIR d'avoir de petite amie ces temps-ci, mais il s'est rendu à de nombreuses reprises à New York dernièrement'. Des connecteurs logiques comme 'si' et des connecteurs discursifs comme 'mais', qui suggèrent des directions d'inférence, sont utilisés par la locutrice pour aider le destinataire à arriver à la conclusion souhaitée (Blakemore 1987, 2002).

Dans les deux scénarios, lorsque Barbara s'exprime de cette façon et qu'André saisit la force de son argument implicite, ils utilisent un mécanisme inférentiel sensible aux relations logiques et évidentielles entre propositions, un mécanisme qui reconnaît, plus spécifiquement, que certaines propositions fonctionnent comme des prémisses et d'autres comme des conclusions. Ce que Barbara communique, et ce qu'André est susceptible de reconnaître, c'est qu'il serait incohérent d'accepter les prémisses et de rejeter la conclusion.

L'argumentation, dans sa forme simple et largement implicite, tout comme dans sa forme plus complexe et explicite, telles que nous les avons illustrées dans les scénarios précédents, est le produit du raisonnement.<sup>7</sup> Dans une série d'articles (Mercier & Sperber 2009, Sperber & Mercier 2012, Sperber 2001, Mercier & Sperber 2011), Hugo Mercier et Dan Sperber ont défendu l'idée que le raisonnement est un outil utilisé dans la vigilance épistémique et dans la communication avec des destinataires vigilants. Sa fonction principale est de permettre aux émetteurs de produire des arguments destinés à convaincre autrui, et aux destinataires d'évaluer des arguments de manière à être convaincus seulement lorsque cela est approprié.

7 Le terme 'raisonnement' est parfois utilisé dans un sens large comme un synonyme d'inférence' (en particulier en psychologie comparative et du développement). Ici nous emploierons 'raisonnement' dans son sens le plus fréquent référant à une forme d'inférence qui suppose la prise en considération des raisons qui nous font accepter une conclusion donnée. Le fait de raisonner, compris de cette façon, implique la réflexion et s'oppose à des formes intuitives d'inférence dans lesquelles on arrive à une conclusion sans considérer les raisons pour lesquelles on l'accepte. Un contraste similaire entre des formes intuitives et réflexives d'inférence a fait l'objet de nombreuses discussions dans les théories du raisonnement en termes de 'processus duels' (voir, par exemple, Evans & Frankish 2009).

Traditionnellement, le raisonnement est considéré comme un outil de la cognition individuelle; il est censé aider les humains à dépasser les limites de l'intuition, à acquérir des croyances plus justifiées – en particulier dans les domaines qui sont hors de la portée de la perception et de l'inférence spontanée – et à prendre de bonnes décisions (Evans & Over 1996, Kahneman 2003, Stanovich 2004). Cette vision répandue s'accommode difficilement des nombreuses données empiriques montrant que le raisonnement humain n'est pas très efficace pour remplir cette prétendue fonction. Le raisonnement ordinaire échoue dans la résolution de problèmes logiques triviaux (Evans 2002) ou lorsqu'il s'agit de contrecarrer des intuitions qui sont fondamentalement défectueuses (Denes-Raj & Epstein 1994). Il nous amène souvent à prendre de mauvaises décisions (Shafir *et al.* 1993, Dijksterhuis *et al.* 2006) et génère des résultats épistémiques médiocres (Kunda 1990). En revanche, l'intuition réussit bien, même dans des computations très complexes (par exemple Trommershauser *et al.* 2008, Balci *et al.* 2009). La prétendue fonction du raisonnement humain comme outil de cognition individuelle ne permet pas d'expliquer ses défauts évidents et un coût opératoire relativement élevé.

Les prédictions qui découlent de la 'théorie argumentative du raisonnement', en revanche, sont abondamment étayées par les travaux effectués dans différents domaines de la psychologie (Mercier 2011a, 2011b, 2012, Mercier & Landemore 2012, Mercier & Sperber 2011). Pour ne donner qu'un exemple, la théorie argumentative fait une prédiction qui la différencie des autres perspectives et qui est particulièrement pertinente pour l'étude de la vigilance épistémique: si la fonction du raisonnement est de trouver des arguments pour convaincre autrui, alors les arguments qu'il produit devraient soutenir la position de l'émetteur et, dans les circonstances appropriées, remettre en cause la position du récepteur. En d'autres termes, le raisonnement devrait exhiber un fort biais de confirmation. Et effectivement, ce biais consistant à «chercher ou interpréter des preuves de manière biaisée en faveur des croyances et des attentes existantes ou d'une hypothèse préalable» (Nickerson 1998: 175) a été documenté dans d'innombrables expériences de psychologie, ainsi que par des observations en milieu naturel (voir Nickerson 1998 pour un passage en revue). Pour les modèles classiques du raisonnement, le fait qu'un tel biais soit aussi prévalent et robuste – ou même le seul fait que ce biais existe – est déroutant, et des limites cognitives

ou motivationnelles ont été invoquées pour tenter de résoudre ce problème. Il y a cependant plusieurs arguments contre ces tentatives de résolution. D'abord, le biais de confirmation semble restreint au raisonnement et ne pas se produire dans les jugements intuitifs (Mercier & Sperber 2011). Ensuite, les tentatives de dépassement de ces prétendus problèmes cognitifs ou motivationnels sont à peu près sans effet (Camerer & Hogarth 1999, Willingham 2008). Cela suggère que le biais de confirmation n'est pas un défaut du raisonnement, mais plutôt une propriété prévisible d'un mécanisme conçu pour persuader autrui par l'usage d'arguments.

Il est intéressant de noter que le biais de confirmation n'engendre pas nécessairement de performances médiocres d'un point de vue logique et normatif. Lorsque des personnes avec des points de vue différents partagent un intérêt authentique dans l'obtention d'une conclusion correcte, le biais de confirmation rend possible une division efficace du travail cognitif. Chaque individu recherche uniquement des raisons soutenant sa propre position, tout en exerçant une vigilance sur les arguments proposés par les autres et en les évaluant avec attention. Cela requiert bien moins de travail qu'une recherche exhaustive des arguments pour et contre chaque position présente dans le groupe. À l'inverse, lorsque les effets du biais de confirmation ne sont pas contrôlés par des personnes d'avis différent, le raisonnement devient risqué du point de vue épistémique et peut rendre les individus sur-confiants quant à leurs propres croyances (Koriat *et al.* 1980) ou les amener à adopter des versions plus extrêmes de ces croyances (Tesser 1978). Dans des discussions de groupe dans lesquelles tous les participants partagent le même point de vue et débattent non pas les uns contre les autres, mais plutôt contre des adversaires absents, un effet de polarisation, susceptible de mener au fanatisme (Sunstein 2002) est communément observé.

Nous ne disons pas que le raisonnement a lieu uniquement dans des contextes communicatifs. Il est clair qu'il se produit également dans la pensée solitaire et qu'il joue un rôle important dans la révision de croyances. Nous nous risquons cependant à spéculer que le raisonnement dans des contextes non communicatifs est une extension d'une composante basique de la capacité de vigilance épistémique envers l'information communiquée, et qu'il implique typiquement un contexte communicatif anticipé ou imaginé. Dans cette perspective, le penseur solitaire est en fait celui qui considère des affirmations auxquelles il pourrait être confronté ou dont il pourrait vouloir

convaincre autrui, ou celui qui engage un dialogue avec lui-même, dialogue dans lequel il alterne entre différents points de vue. Les données expérimentales pourront aider à confirmer ou à infirmer cette hypothèse: par exemple, nous prédisons que le fait d'encourager ou d'inhiber l'anticipation ou l'imagination d'un tel contexte faciliterait, ou, au contraire, entraverait le raisonnement.

## 8. Vigilance épistémique à l'échelle d'une population

Nous avons jusqu'à présent considéré le rôle de filtre joué par la vigilance épistémique dans le flux d'informations propre aux interactions en face à face. Dans cette section, nous allons maintenant nous intéresser au flux d'informations à l'échelle d'une population, comme c'est le cas par exemple pour l'émergence de bonnes ou de mauvaises réputations ou de la propagation de croyances religieuses. Des informations de ce type, qui se propagent dans une population par transmission sociale, sont connues sous le nom 'd'informations culturelles'. Le succès social, qui est pour ainsi dire une propriété définitoire de l'information culturelle, pourrait suggérer que celle-ci est acceptée sans aucun contrôle critique (mis à part dans des cas de conflit culturel). Nous défendrons cependant l'idée que dans ces cas la vigilance épistémique est également à l'œuvre, bien qu'elle requière des développements culturels et institutionnels appropriés pour relever les défis épistémiques posés par l'information culturelle.

Aucun acte de communication entre humains, même si sa pertinence pour les interlocuteurs au moment où il a lieu n'est que locale, n'est totalement déconnecté du flux d'informations du groupe social dans son entier. La communication humaine a toujours recours à des éléments culturels. Ces recours peuvent être explicites, comme lorsque André dit à Barbara 'Le champagne, c'est cher', en lui rappelant une hypothèse culturellement partagée dans leur milieu et qu'il considère pertinente dans les circonstances. Ces recours peuvent être implicites, comme lorsqu'André dit à Barbara, dans un autre de nos exemples, 'Steve n'a pas l'air d'avoir de petite amie en ce moment'. Même si la remarque d'André a une pertinence locale et assez

limitée, elle implicite des hypothèses culturellement partagées sur ce que l'on peut attendre d'un célibataire comme Steve, sans quoi sa remarque ne serait pas pertinente au sens où il le désirait. De nombreuses hypothèses culturelles sont distribuées de cette façon, pas tant – voire, dans certains cas, pas du tout – en étant directement assertées, mais en étant utilisées comme des prémisses implicites dans un grand nombre d'actes communicatifs tels que l'énoncé d'André.

Exerçons-nous notre vigilance épistémique dans le cas de contenus culturellement transmis, et si oui, comment? Lorsque de tels contenus sont communiqués, explicitement ou implicitement, les émetteurs utilisent leur propre autorité individuelle pas tant pour endosser le contenu en question que pour se porter garants de son statut d'hypothèse culturelle communément acceptée. Lorsqu'André dit que le champagne est cher, ou implicite qu'il serait normal que Steve ait une petite amie, il communique que ces contenus sont acceptés dans son milieu et celui de Barbara. Si Barbara n'est pas d'accord, elle manifeste non seulement un désaccord envers André, mais également un désaccord quant à une vision communément acceptée.

Si une idée est généralement acceptée par les personnes avec lesquelles vous interagissez, n'est-ce pas là une bonne raison pour vous de l'accepter également? Il peut s'avérer modeste et prudent de se rallier aux avis des gens avec qui l'on interagit et d'accepter les idées qu'ils acceptent. N'importe quelle autre attitude pourrait compromettre la compétence culturelle et l'acceptabilité sociale d'une personne. Il est tout à fait possible que les membres d'un groupe culturel acceptent ce qu'ils prennent pour des 'opinions acceptées' dans ce sens pragmatique, sans devoir s'engager de manière forte ou claire sur leur contenu (Sperber 1975, 1985, Boyer 1990, Bloch 1998). D'un point de vue épistémologique, le fait qu'une idée soit largement partagée n'est pas une bonne raison de l'accepter, à moins que ceux qui l'acceptent ne l'aient acceptée indépendamment les uns des autres. C'est uniquement dans ces circonstances que le fait de savoir que chaque individu accepte cette idée pourra jouer un rôle dans nos propres raisons épistémologiques de l'adopter à notre tour. Très souvent, cependant, les gens qui acceptent (dans un sens épistémologique) des idées culturelles partagées n'ont pas de raisons indépendantes de le faire.

Souvent, l'information se propage à travers un groupe depuis une source unique et est acceptée par ceux qui se trouvent tout au long de la chaîne de

transmission parce qu'ils font confiance à cette source plus que parce qu'ils disposent de preuves ou d'arguments en faveur du contenu. Si c'est le cas, la considération cruciale devrait être la crédibilité de la source originelle. Si chaque personne qui relaie l'information a de bonnes raisons indépendantes de faire confiance à la source, cela devrait donner aux personnes se trouvant en aval dans la chaîne de bonnes raisons de faire également confiance à la source, et par conséquent d'accepter le contenu initialement communiqué. Cependant, les raisons de croire la source ne sont généralement pas plus indépendantes les unes des autres que les raisons qu'ils ont d'accepter le contenu.

Même si nous faisons l'hypothèse forte que chaque individu situé sur la chaîne de transmission de l'information entre la source et nous-mêmes a de bonnes raisons de faire confiance à l'individu précédent dans la chaîne, ces raisons peuvent toujours être erronées; par conséquent, notre propre confiance dans la source originelle devrait diminuer à mesure que la longueur de la chaîne s'accroît. De plus, il est assez courant qu'une information dont la source n'est pas clairement identifiée soit acceptée et transmise uniquement sur la base du fait qu'elle est largement acceptée et transmise – ce qui constitue une circularité évidente. Ajoutons à ces considérations le fait que lorsqu'une idée se propage à travers une population, son contenu tend à être altéré durant le processus sans que les propagateurs n'en soient conscients (comme c'est le cas de pratiquement toutes les rumeurs et traditions – voir Sperber 1996). Dans ces cas, même s'il existait de bonnes raisons de considérer que la source originelle est crédible, cela ne fournirait aucun soutien sérieux à l'idée telle qu'elle est formulée aujourd'hui.

Dès lors, il semblerait que les humains sont simplement disposés, ou même désireux, d'accepter de l'information culturellement transmise sans exercer de vigilance épistémique à son endroit. Boyd, Richerson et Henrich ont défendu l'idée de l'évolution d'un biais psychologique conformiste en faveur de l'adoption du comportement et des croyances de la majorité des membres de sa communauté (par ex. Boyd & Richerson 1985, Henrich & Boyd 1998). Csibra & Gergely (2009) ont soutenu que les humains en général, et les enfants en particulier, sont demandeurs d'informations culturelles, et que cela pourrait les biaiser dans l'interprétation (et même la surinterprétation) de la pertinence culturelle de l'information communiquée, tout comme dans son acceptation. Une hypothèse alternative (ou peut-être



complémentaire) voudrait que les humains exercent un degré de vigilance épistémique envers toute information communiquée, qu'elle soit locale ou culturelle, mais que leur vigilance vise primordialement l'information issue d'interactions en face à face, et non les informations propagées à plus large échelle. Par exemple, les humains pourraient être disposés à prêter attention aux problèmes soulevés par la non-indépendance des témoignages ou par des problèmes de consistance dans leur contenu lorsque ceux-ci sont flagrants, comme c'est le cas lorsqu'ils sont détectés dans des situations d'interaction en face à face, mais pas dans d'autres circonstances. A l'échelle d'une population, ces problèmes peuvent passer inaperçus même si, à la réflexion, ils sont susceptibles d'être largement répandus. Toutes sortes de croyances largement partagées dans une communauté sont susceptibles de se propager au sein d'une culture en faisant appel à la confiance individuelle en des témoignages convergents. La confiance n'est pas aveugle, mais la vigilance épistémique censée l'étayer est myope.

Il est particulièrement pertinent de considérer ici deux types de croyances typiquement culturelles: les réputations et les croyances dont le contenu est mystérieux et qui donc ne sont que partiellement comprises. Il existe une littérature abondante sur la réputation en général (par ex. Tirole 1996, Morris 1999) et sur sa pertinence vis-à-vis de la coopération et de l'épistémologie sociale (par ex. Mathew & Boyd 2009).

Le terme de 'réputation' est généralement compris comme dénotant une opinion positive ou négative, par exemple l'opinion selon laquelle Lisa est généreuse ou selon laquelle Jean est un menteur, opinion devenue largement acceptée dans un groupe par transmission répétée. Si un individu appartient à un groupe relativement restreint dans lequel plusieurs personnes ont une expérience directe des qualités et des défauts de Lisa, et dans lequel ces personnes peuvent exprimer et comparer leurs opinions avec une certaine liberté (par exemple par le commérage), alors le comportement de Lisa peut jouer un rôle important dans le renforcement ou dans la détérioration de sa réputation. Bien sûr, les personnes qui commèrent peuvent elles-mêmes être incompetentes ou pas tout à fait honnêtes, mais la vigilance épistémique ordinaire est pertinente pour l'évaluation à la fois des commérages et de ceux qui les propagent. Cependant, de nombreuses réputations sont propagées à plus large échelle, par des gens sans connaissances pertinentes pour l'évaluation directe de ces réputations. Lorsqu'un destinataire doit décider si

oui ou non il croit une source d'information non familière, il peut se trouver dans la situation où il n'a aucune autre base pour prendre sa décision que sa connaissance de la réputation de la source, qu'il ne peut évaluer lui-même, et qu'il est susceptible d'accepter à défaut d'un meilleur choix. Bien trop souvent, les réputations sont des exemples d'idées qui sont acceptées et transmises purement sur la base du fait qu'elles sont largement acceptées et transmises.

Comme nous l'avons remarqué précédemment, le contenu de croyances socialement transmises est typiquement modifié au cours de leur transmission. L'une des transformations responsables de la modification des réputations est le fait qu'elles gonflent au-delà du niveau de celui des opinions typiques auxquelles on arrive individuellement. Lorsque des autorités épistémiques – leaders religieux, gourous, maîtres à penser – obtiennent de telles réputations artificiellement élevées, les gens qui sont dès lors plus enclins à s'en remettre à ces autorités qu'à des sources dont ils ont directement évalué la crédibilité peuvent se trouver dans le dilemme suivant: s'ils devaient évaluer les déclarations de ces sources (par exemple 'Marie était et est restée vierge lorsqu'elle a accouché' ou la déclaration de Lacan selon laquelle 'il n'y a pas de rapport sexuel') en termes de cohérence vis-à-vis de leurs croyances existantes, ils les rejetteraient. Mais cela les amènerait ensuite à remettre en question leur acceptation de l'autorité de la source. Une solution répandue à ce problème consiste à engager une variante de 'l'interprétation charitable' davidsonienne, et 'd'optimiser l'accord' non pas en fournissant une interprétation claire et acceptable de ces déclarations, mais en s'en remettant aux autorités (ou aux interprètes autorisés) pour l'interprétation appropriée, et par conséquent d'accepter une idée seulement comprise à moitié ou 'semi-propositionnelle' (Sperber 1985, 1997, 2010). La plupart des croyances religieuses sont des exemples typiques de ce genre de croyance, dont le contenu est en partie mystérieux pour les croyants eux-mêmes (Bloch 1998, Boyer 2001).

Pour l'instant, l'image que nous avons esquissée de la vigilance épistémique à l'échelle d'une population est peu réjouissante. Les mécanismes de vigilance épistémique ne sont pas calibrés pour le filtrage d'informations transmises à si grande échelle. Même si nous avons raison de considérer que ces mécanismes existent, ils n'empêchent pas des idées erronées, des réputations non méritées ou des croyances creuses d'envahir

des populations entières. En revanche, nous avons fait remarquer qu'il est important de ne pas conclure du fait que les gens sont sérieusement, même passionnément, engagés sur certaines idées et qu'ils attendent des autres qu'ils soient similairement engagés, que l'engagement en question est clairement épistémique. Il peut arriver que le contenu des idées vous importe moins que de savoir avec qui vous les partagez dans la mesure où des idées partagées peuvent aider à définir des identités de groupe. Lorsque ce qui importe est le partage, des contenus qui se prêtent sans aucun problème à une évaluation épistémique risquent soit de soulever des objections dans le groupe social visé, soit d'être trop facilement partagés au-delà de ce groupe. Par conséquent, il est possible que des contenus semi-propositionnels qui sont acceptés sans aucun problème par le groupe en question jouissent d'un succès culturel inversement proportionnel à leur valeur épistémique.

Jusqu'à présent, nous avons seulement considéré les effets de la vigilance ordinaire individuelle exercée à l'échelle d'une population. Cependant, la vigilance épistémique peut prendre une forme institutionnelle. Certaines de ces institutions aident à protéger les autorités établies ou à imposer un dogme, et sont par conséquent préjudiciables à de véritables buts épistémiques. De telles formes de vigilance hégémonique ou dogmatique ne sont pas *stricto sensu* épistémiques. A l'inverse, d'autres formes institutionnelles de vigilance épistémique authentique peuvent fournir un meilleur filtrage épistémique que le seul effet cumulatif de la vigilance spontanée exercée par les individus.

Dans un certain nombre de domaines, il existe des procédures institutionnelles qui ont pour objectif d'évaluer la compétence des individus, qui rendent ces évaluations publiques, au moyen d'une forme de certification, et qui sanctionnent toute prétention frauduleuse à être ainsi certifié. Les docteurs, professeurs, juges, géomètres, comptables, prêtres, et bien d'autres encore, sont généralement tenus pour experts dans leur domaine parce qu'ils ont fourni des preuves solides de leur expertise à des experts encore mieux qualifiés qu'eux. Bien sûr, ces procédures peuvent s'avérer inadéquates ou corrompues, et le domaine peut lui-même être criblé d'erreurs; il n'en reste pas moins que de telles procédures fournissent des indications claires et aisément accessibles quant à l'expertise d'un individu.

Lorsqu'on analyse la façon dont l'information est évaluée, filtrée et, par la même occasion, transformée à l'échelle de la population, il est important

d'étudier la vigilance envers le contenu tout autant que la vigilance envers la source. Nous avons suggéré plus haut que la vigilance envers le contenu est typiquement exercée à travers le débat et l'argumentation, et qu'elle peut donner lieu à une sorte de division spontanée du travail cognitif. Une telle division du travail peut elle-même être organisée culturellement et prendre des formes institutionnelles variées. Les institutions judiciaires fournissent de tels exemples à travers, entre autres, la conception de nombreuses règles et procédures destinées à établir les faits par l'examen des preuves, l'interrogation de témoins et les débats entre parties adverses. L'organisation institutionnelle de la vigilance épistémique n'est nulle part plus évidente que dans les sciences, au sein desquelles les hypothèses fondées sur des observations ou des théories sont évaluées de façon critique via des processus sociaux comme la discussion en laboratoire, les ateliers de recherche, les conférences et l'examen par les pairs dans des revues à comité de lecture. La fiabilité d'une revue est elle-même évaluée par des classements, et ainsi de suite (Goldman 1999).

Les mécanismes sociaux de vigilance envers la source interagissent avec ceux qui portent sur le contenu de plusieurs manières. Dans les procédures judiciaires, par exemple, la réputation des témoins est minutieusement examinée de façon à renforcer ou affaiblir leur témoignage. Dans les sciences, l'examen par les pairs est conçu pour être exclusivement orienté vers le contenu, mais il est trop souvent influencé par la réputation préalable des auteurs (même si l'évaluation à l'aveugle est censée supprimer cette influence), et le résultat de cet examen affecte à son tour la réputation de l'auteur. La certification d'une expertise, par exemple l'octroi d'un doctorat, implique généralement des évaluations complexes et multiples de la part d'enseignants et d'examineurs au cours d'échanges avec le candidat, ainsi qu'entre les experts; ces évaluations sont compilées par des institutions éducatives qui délivrent finalement un label de réputation, le 'doctorat', destiné à un usage public.

Nous ne pouvons ici faire guère plus que de relever certains des problèmes soulevés par les mécanismes sociaux de vigilance épistémique. Notre objectif principal dans cette démarche est de suggérer que ces mécanismes sociaux, dans une large mesure, sont l'articulation de mécanismes psychologiques liés par des chaînes de communication et, dans ces certains cas, par un modelage institutionnel (Sperber 1996). Dans ces

articulations à l'échelle de la population, les mécanismes psychologiques se combinent avec des artefacts cognitifs (par exemple des instruments de mesure), des techniques (comme des tests statistiques de confiance) et des procédures (le contre-interrogatoire) pour générer des systèmes distribués d'évaluation épistémique (Heintz 2006) qui doivent être considérés comme un type spécial de système cognitif distribué (Hutchins 1996).

La façon dont les humains s'appuient sur des systèmes distribués d'évaluation pose une nouvelle version du problème de Reid et de Hume sur la façon dont on peut justifier notre confiance dans le témoignage. Cela est particulièrement vrai dans le cas des nouveaux systèmes d'évaluation sans lesquels nous serions incapables d'utiliser Internet. Google en est un exemple saillant. Google n'est pas seulement un moteur de recherche: il est aussi utilisé comme un moteur d'évaluation épistémique. Il représente de manière implicite, sous la forme d'un classement, la valeur épistémique relative des documents trouvés en ligne à la suite d'une recherche. Plus la position du document est élevée dans le classement, plus celui-ci est susceptible de contenir des informations pertinentes et crédibles. L'une des méthodes de production d'un tel classement implique le calcul du nombre de liens, vers un document donné, qui proviennent d'autres sites web, et la pondération de ce nombre par rapport à l'importance relative de ces sites (qui est elle-même calculée sur la base du nombre de liens entrants en provenance d'autres sites). L'idée générale qui sous-tend ce processus est que le fait de créer un lien vers un document est un jugement implicite de sa valeur. Le processus compile ces jugements dans une indication accessible: sa position dans un classement sur une page de résultats.

Pourquoi, dès lors, les gens se fient-ils à un moteur de recherche comme Google même s'ils ne savent que très peu, voire rien, de la façon dont il génère ses résultats? Et à quel point sont-ils justifiés à le faire? Il convient de remarquer, d'abord, que notre confiance n'est pas entièrement aveugle: cette technologie cognitive, comme n'importe quelle autre technologie, est adoptée parce que son succès à déjà été observé. De plus, notre confiance est provisoire: nous sommes disposés d'abord à regarder les pages figurant en tête de classement et à faire l'hypothèse qu'il y a de bonnes raisons pour qu'elles y figurent. Mais n'exerçons-nous pas alors une forme typique de vigilance épistémique envers l'information qui nous est présentée? Les

travaux et les idées évoquées dans cet article devraient s'avérer pertinents pour une investigation empirique de ces nouvelles questions.

## 9. Remarque conclusive

Notre but dans cet article était d'étayer la thèse selon laquelle les humains disposent d'une série de mécanismes cognitifs de vigilance épistémique. A cette fin, nous avons passé en revue des problématiques, des recherches et des théories dans différents domaines de la philosophie, de la linguistique, de la psychologie cognitive et des sciences sociales. Nous ne nous attendons pas à ce que nos lecteurs aient accepté toutes nos hypothèses, dont plusieurs nous paraissent d'ailleurs plutôt spéculatives. Nous espérons toutefois avoir réussi à mettre en avant l'importance de la vigilance épistémique dans les interactions humaines. Tout comme la communication, à laquelle elle est essentiellement liée, la vigilance épistémique s'appuie sur des mécanismes mentaux individuels qui s'articulent entre individus et au sein de populations dans des mécanismes sociaux. Certains de ces mécanismes sont dirigés vers la source de l'information, d'autres vers son contenu. Considérer ces divers mécanismes comme autant de contributions à une seule et même fonction de vigilance épistémique peut servir à éclairer l'étude de chacun de ces mécanismes pris séparément.

## Bibliographie

- Adler, J. (2002): *Belief's Own Ethics*. Cambridge, MA: MIT Press.
- Aikhenvald, A. Y. (2004): *Evidentiality*. Oxford: Oxford University Press.
- Ainslie, G. (2001): *Breakdown of Will*. Cambridge: Cambridge University Press.
- Alexander, R. D. (1987): *The Biology of Moral Systems*. Berlin: Aldine de Gruyter.
- Asch, S. E. (1956): «Studies of independence and conformity: a minority of one against a unanimous majority», *Psychological Monographs* 70, 1-70.
- Austin, J. L. (1962): *How to Do Things with Words*. Cambridge, MA: Harvard University Press.

- Balci, F., Freestone, D. & Gallistel, C. R. (2009): «Risk assessment in man and mouse», *PNAS* 106, 2459-2463.
- Baron-Cohen, S., Leslie, A. M. & Frith, U. (1985): «Does the autistic child have a ‘theory of mind?’», *Cognition* 21, 37-46.
- Bergstrom, B. & Boyer, P. (ms.): «Who mental systems believe: effects of source on judgments of truth».
- Bergstrom, B., Moehlmann, B. & Boyer, P. (2006): «Extending the testimony problem: Evaluating the truth, scope, and source of cultural information», *Child Development* 77, 531-538.
- Bezuidenhout, A. (1998): «Is verbal communication a purely preservative process?», *The Philosophical Review* 107, 261-288.
- Birch, S., Vauthier, S. A. & Bloom, P. (2008): «Three- and four-year-olds spontaneously use others’ past performance to guide their learning», *Cognition*, 107, 1018-1034.
- Birch, S. A. J., Akmal, N. & Frampton, K. L. (2009): «Two-year-olds are vigilant of others’ non-verbal cues to credibility». *Developmental Science* 13, 363-369.
- Blakemore, D. (1987): *Semantic Constraints on Relevance*. Oxford: Blackwell.
- Blakemore, D. (2002): *Relevance and Linguistic Meaning: The Semantics and Pragmatics of Discourse Markers*. Cambridge: Cambridge University Press.
- Bloch, M. (1998): *How We Think They Think: Anthropological Approaches to Cognition, Memory, and Literacy*. Boulder, CO: Westview Press.
- Bond, C. F. Jr. & DePaulo, B. M. (2006): «Accuracy of deception judgments», *Personality and Social Psychology Review* 10, 214-234.
- Boyer, P. (1992): *Tradition as Truth and Communication*. Cambridge: Cambridge University Press.
- Boyer, P. (2001): *Religion Explained: The Evolutionary Foundations of Religious Belief*. New York: Basic Books.
- Boyd, R. & Richerson, P. J. (1985): *Culture and the Evolutionary Process*. Chicago, IL: Chicago University Press.
- Burge, T. (1993): «Content preservation», *Philosophical Review* 101, 457-488.
- Call, J. & Tomasello, M. (1999): «A nonverbal false belief task: the performance of children and great apes», *Child Development* 70, 381-395.
- Camerer, C. & Hogarth, R. M. (1999): «The effect of financial incentives on performance in experiments: a review and capital-labor theory», *Journal of Risk and Uncertainty* 19, 7-42.
- Carlson, S. M. & Moses, L. J. (2001): «Individual differences in inhibitory control and children’s theory of mind», *Child Development* 72, 1032-1053.
- Carston, R. (2002): *Thoughts and Utterances: The Pragmatics of Explicit Communication*. Oxford: Blackwell.
- Chafe, W. & Nichols, J. eds. (1986): *Evidentiality: The Linguistic Coding of Epistemology*. Norwood, NJ: Ablex.
- Chaiken, S. (1980): «Heuristic versus systematic information processing and the use of source versus message cues in persuasion», *Journal of Personality and Social Psychology* 39, 752-766.
- Cheney, D. L. & Seyfarth, R. M. (1990): *How Monkeys See the World*. Chicago, IL: Chicago University Press.
- Clément, F. (2010): «To trust or not to trust? Children’s social epistemology», *Review of Philosophy and Psychology* 1(4): 531-549.

- Clément, F., Koenig, M. A. & Harris, P. (2004): «The ontogeny of trust», *Mind & Language* 19, 360-379.
- Van Cleve, J. (2006): «Reid on the Credit of Human Testimony», in J. Lackey & E. Sosa (eds), *The Epistemology of Testimony*. Oxford: Oxford University Press.
- Coady, C. A. J. (1992): *Testimony*. Oxford: Oxford University Press.
- Corriveau, K. & Harris, P. L. (2009): «Preschoolers continue to trust a more accurate informant 1 week after exposure to accuracy information», *Developmental Science* 12, 188-193.
- Corriveau, K. H. & Harris, P. L. (2010): «Young children's trust in what other people say», in Rotenberg, K. (ed.), *Interpersonal Trust During Childhood and Adolescence*. Cambridge: Cambridge University Press.
- Cosmides, L. & Tooby, J. (2005): «Neurocognitive adaptations designed for social exchange», in D. M. Buss (ed.), *Evolutionary Psychology Handbook*. New York: Wiley, 584-627.
- Couillard, N. L. & Woodward, A. L. (1999): «Children's comprehension of deceptive points», *British Journal of Developmental Psychology* 17, 515-521.
- Csibra, G. & Gergely, G. (2009): «Natural pedagogy», *Trends in Cognitive Sciences* 13, 148-153.
- Davidson, D. (1984): «Radical interpretation», in his *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press, 125-140.
- Davidson, D. (1986): «A coherence theory of truth and knowledge», in E. LePore (ed), *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*. Oxford: Basil Blackwell, 307-319.
- Dawkins, R. & Krebs, J. R. (1978): «Animal signals: information or manipulation?», in Krebs, J. R. & Davies, N. B. (eds), *Behavioural Ecology: An Evolutionary Approach*. Oxford: Basil Blackwell Scientific Publications.
- Denes-Raj, V. & Epstein, S. (1994): «Conflict between intuitive and rational processing: when people behave against their better judgment», *Journal of Personality and Social Psychology* 66, 819-829.
- DePaulo, B. M. & Pfeifer, R. L. (1986): «On-the-job experience and skill at detecting deception», *Journal of Applied Social Psychology* 16, 249-267.
- Dijksterhuis, A., Bos, M. W., Nordgren, L. F. & Van Baaren, R. B. (2006): «On making the right choice: the deliberation-without-attention effect», *Science* 311, 1005-1007.
- Ekman, P. (2001): *Telling Lies*. New York: Norton.
- Ekman, P. & O'Sullivan, M. (1991): «Who can catch a liar», *American Psychologist* 46, 913-920.
- Evans, J. S. B. T. (2002): «Logic and human reasoning: an assessment of the deduction paradigm», *Psychological Bulletin* 128, 978-996.
- Evans, J. S. B. T. & Frankish, K. (eds) (2009): *In Two Minds*. Oxford: Oxford University Press.
- Evans, J. S. B. T. & Over, D. E. (1996): *Rationality and Reasoning*. Hove: Psychology Press.
- Figueras-Costa, B. & Harris, P. (2001): «Theory of mind development in deaf children: a nonverbal test of false-belief understanding», *Journal of Deaf Studies and Deaf Education* 6, 92.
- Foley, R. (1994): «Egoism in epistemology», in Schmitt, F. (ed), *Socializing Epistemology*. Lanham, MD: Rowman & Littlefield, Inc., 53-73.
- Fricker, E. (1995): «Critical notice: telling and trusting: reductionism and anti-reductionism in the epistemology of testimony», *Mind* 104, 393-411.



- Fricker, E. (2006): «Testimony and epistemic autonomy», in Lackey, J. & Sosa, E. (eds), *The Epistemology of Testimony*. Oxford: Oxford University Press.
- Fusaro, M. & Harris, P. L. (2008): «Children assess informant reliability using bystanders' non-verbal cues», *Developmental Science* 11, 771-777.
- Gee, C. L. & Heyman, G. D. (2007): «Children's evaluation of other people's self-descriptions», *Social Development* 16, 800-818.
- Gilbert, D. T., Krull, D. S. & Malone, P. S. (1990): «Unbelieving the unbelievable: some problems in the rejection of false information», *Journal of Personality and Social Psychology* 59, 601-613.
- Gilbert, D. T. & Malone, P. S. (1995): «The correspondence bias», *Psychological Bulletin* 117, 21-38.
- Gilbert, D. T., Tafarodi, R. W. & Malone, P. S. (1993): «You can't not believe everything you read», *Journal of Personality and Social Psychology* 65, 221-233.
- Goldman, A. (1999): *Knowledge in a Social World*. Oxford: Oxford University Press.
- Gouzoules, H., Gouzoules, S. & Miller, K. (1996): «Skeptical responding in rhesus monkeys (Macaca mulatta)», *International Journal of Primatology* 17, 549-568.
- Grice, H. P. (1957): «Meaning», *Philosophical Review* 66, 377-388.
- Grice, H.P. (1969): «Utterer's meaning and intentions», *Philosophical Review* 78, 147-177. [Réimprimé in Grice, (1989, 86-116)].
- Grice, H. P. (1975): «Logic and conversation», in Cole, P. & Morgan, J. P. (eds), *Syntax and Semantics* Vol. 3: Speech Acts. New York: Seminar Press.
- Grice, H. P. (1989): *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.
- Hardwig, J. (1985): «Epistemic dependence», *The Journal of Philosophy* 82, 335-349.
- Harman, G. (1986): *Change in View: Principles of Reasoning*. Cambridge, MA: MIT Press.
- Harris, P. & Núñez, M. (1996): «Understanding of permission rules by preschool children», *Child Development* 67, 1572-1591.
- Hasson, U., Simmons, J. P. & Todorov, A. (2005): «Believe it or not: on the possibility of suspending belief», *Psychological Science* 16, 566-571.
- Heintz, C. (2006): «Web search engines and distributed assessment systems», *Pragmatics & Cognition* 14, 387-409.
- Henrich, J. & Boyd, R. (1998): «The evolution of conformist transmission and the emergence of between-group differences», *Evolution and Human Behavior* 19, 215-241.
- Heyman, G. D. (2008): «Children's critical thinking when learning from others», *Current Directions in Psychological Science* 17, 344-347.
- Holton, R. (1994): «Deciding to trust, coming to believe», *Australasian Journal of Philosophy* 72, 63-76.
- Hutchins, E. (1980): *Culture and Inference: A Trobriand Case Study*. Cambridge, MA: Harvard University Press.
- Hutchins, E. (1996): *Cognition in the Wild*. Cambridge, MA: MIT Press.
- Ifantidou, E. (2001): *Evidentials and Relevance*. Amsterdam: John Benjamins.
- Jaswal, V. K., Croft, A. C., Setia, A. R. & Cole, C. A. (2010): «Young children have a specific, highly robust bias to trust testimony», *Psychological Science* 21: 1541-1547.
- Kahneman, D. (2003): «A perspective on judgment and choice: mapping bounded rationality», *American Psychologist* 58, 697-720.

- Koenig, M. A. & Echols, C. H. (2003): «Infants' understanding of false labeling events: the referential roles of words and the speakers who use them», *Cognition* 87, 179-203.
- Koenig, M. A. & Harris, P. L. (2007): «The basis of epistemic trust: reliable testimony or reliable sources?», *Episteme* 4, 264-284.
- Koriat, A., Lichtenstein, S. & Fischhoff, B. (1980): «Reasons for confidence», *Journal of Experimental Psychology: Human Learning and Memory and Cognition* 6, 107-118.
- Krebs, J. R. & Dawkins, R. (1984): «Animal signals: mind-reading and manipulation?», in Krebs, J. R. & Davies, N. B. (eds), *Behavioural Ecology: An Evolutionary Approach*, 2nd edn. Oxford: Basil Blackwell Scientific Publications.
- Kunda, Z. (1990): «The case for motivated reasoning», *Psychological Bulletin* 108, 480-498.
- Lampinen, J. M. & Smith, V. L. (1995): «The incredible (and sometimes incredulous) child witness: child eyewitnesses' sensitivity to source credibility cues», *Journal of Applied Psychology* 80, 621-627.
- Lewis, D. K. (1969): *Conventions*. Cambridge, MA: Harvard University Press.
- Lewis, M., Stranger, C. & Sullivan, M. W. (1989): «Deception in 3-year-olds», *Developmental Psychology* 25, 439-443.
- Locke, J. (1975) [1690]: *An Essay Concerning Human Understanding*, P. Nidditch, ed. Oxford: Oxford University Press.
- Lutz, D. J. & Keil, F. C. (2002): «Early understanding of the division of cognitive labor», *Child Development* 1073-1084.
- Malone, B. E. & DePaulo, B. M. (2001): «Measuring sensitivity to deception», in Hall, J. A. & Bernieri, F. (eds), *Interpersonal Sensitivity: Theory, Measurement, and Application*. Hillsdale, NJ: Erlbaum, 103-124.
- Mann, S., Vrij, A. & Bull, R. (2004): «Detecting true lies: police officers' ability to detect suspects' lies», *Journal of Applied Psychology* 89, 137-149.
- Mascaro, O. & Sperber, D. (2009): «The moral, epistemic, and mindreading components of children's vigilance towards deception», *Cognition* 112, 367-380.
- Mathew, S. & Boyd, R. (2009): «When does optional participation allow the evolution of cooperation?», *Proceedings of the Royal Society of London B* 276(1659), 1167-1174.
- Matsui, T., Rakoczy, H., Miura, Y. & Tomasello, M. (2009): «Understanding of speaker certainty and false-belief reasoning: a comparison of Japanese and German preschoolers», *Developmental Science* 12, 602-613.
- Mercier, H. (2011a). «Reasoning serves argumentation in children», *Cognitive Development* (26)3, 177-191.
- Mercier, H. (2011b): «On the universality of argumentative reasoning», *Journal of cognition and culture* 11(1), 85-113.
- Mercier, H. (2012): «Looking for arguments», *Argumentation* 26(3), 305-324.
- Mercier, H. & Landemore, H. (2012). «Reasoning is for arguing: Understanding the successes and failures of deliberation», *Political Psychology* 33(2), 243-258.
- Mercier, H. & Sperber, D. (2009): «Intuitive and reflective inferences», in Evans, J. St B. T. & Frankish, K. (eds), *In Two Minds*. New York: Oxford University Press, 149-170.
- Mercier, H. & Sperber, D. (2011): «Why do humans reason? Arguments for an argumentative theory», *Behavioral and Brain Science* 34, 57 -111.

- Milinski, M., Semmann, D. & Krambeck, H. J. (2002): «Reputation helps solve the ‘tragedy of the commons’», *Nature* 415, 424-426.
- Millikan, R. G. (1987): *Language, Thought and Other Categories*. Cambridge, MA: MIT Press.
- Mills, C. M. & Keil, F. C. (2005): «The Development of Cynicism», *Psychological Science* 16, 385-390.
- Morris, C.W. (1999): «What is this thing called ‘reputation?’», *Business Ethics Quarterly* 9, 87-102.
- Nickerson, R. S. (1998): «Confirmation bias: a ubiquitous phenomena in many guises», *Review of General Psychology* 2, 175-220.
- Nowak, M. A. & Sigmund, K. (1998): «Evolution of indirect reciprocity by image scoring», *Nature* 383, 573-577.
- Nurmsoo, E. & Robinson, E. J. (2009): «Children’s trust in previously inaccurate informants who were well or poorly informed: when past errors can be excused», *Child Development* 80, 23-27.
- Nurmsoo, E., Robinson, E. J. & Butterfill, S. A. (2010): «Children’s Selective Learning from Others», *Review of Psychology and Philosophy* 1(4), 551-561.
- Onishi, K. H. & Baillargeon, R. (2005): «Do 15-month-old infants understand false beliefs?», *Science* 308, 255-258.
- Origgi, G. (2005): «A stance of trust», Paper presented at the 9th International Pragmatics Conference (IPRA), Riva del Garda, July 10-15th; to be published in T. Matsui (ed.): *Pragmatics and Theory of Mind*. Amsterdam: John Benjamins (forthcoming).
- Origgi, G. (2008): «Trust, authority and epistemic responsibility», *Theoria* 23, 35-44.
- Origgi, G. & Sperber, D. (2000): «Evolution, communication and the proper function of language», in Carruthers P. and Chamberlain, A. (eds), *Evolution and the Human Mind: Modularity, Language and Meta-Cognition*. Cambridge: Cambridge University Press, 140-169.
- Pagliari, F. & Woods, J. (2011): «Enthymematic parsimony», *Synthese* 178(3), 461-501.
- Pea, R. D. (1982): «Origins of verbal logic: spontaneous denials by two-and three-yearolds», *Journal of Child Language* 9, 597-626.
- Perner, J. & Lang, B. (2000): «Theory of mind and executive function: Is there a developmental relationship?», in Baron-Cohen, S., Tager-Flusberg, H. & Cohen, D. (eds), *Understanding Other Minds: Perspectives From Autism and Developmental Cognitive Neuroscience*, 2nd edn. Oxford: Oxford University Press, 150-181.
- Peskin, J. (1996): «Guise and guile: children’s understanding of narratives in which the purpose of pretense is deception», *Child Development* 1735-1751.
- Petty, R. E. & Cacioppo, J. T. (1986): «The elaboration likelihood model of persuasion», in Berkowitz, L. (ed.), *Advances in Experimental Social Psychology*. Orlando, FL: Academic Press, 124-206.
- Povinelli, D. J. & Deblois, S. (1992): «Young children’s (Homo sapiens) understanding of knowledge formation in themselves and others», *Journal of Comparative Psychology* 106, 228-238.
- Reid, T. (2000) [1764]: *Inquiry into the Human Mind*, T. Duggan, ed. Chicago, IL: University of Chicago Press.
- Richter, T., Schroeder, S. & Wöhrmann, B. (2009): «You don’t have to believe everything you read: background knowledge permits fast and efficient validation of information», *Journal of Personality and Social Psychology* 96, 538-558.

- Robinson, E. J., Haigh, S. N. & Nurmsoo, E. (2008): «Children's working understanding of knowledge sources: confidence in knowledge gained from testimony», *Cognitive Development* 23, 105-118.
- Robinson, E. J. & Whitcombe, E. L. (2003): «Children's suggestibility in relation to their understanding about sources of knowledge», *Child Development* 74, 48-62.
- Rosen, L. (1989): *The Anthropology of Justice: Law as Culture in Islamic Society*. Cambridge: Cambridge University Press.
- Ross, L. (1977): «The intuitive psychologist and his shortcomings: distortions in the attribution process», in Berkowitz, L. (ed.), *Advances in Experimental Social Psychology*. New York: Academic Press, 173-220.
- Ross, L. & Nisbett, R. E. (1991): *The Person and the Situation: Perspectives of Social Psychology*. New York: McGraw-Hill.
- Sabbagh, M. A. & Baldwin, D. A. (2001): «Learning words from knowledgeable versus ignorant speakers: links between preschoolers' theory of mind and semantic development», *Child Development* 72, 1054-1070.
- Schiffer, S. (1972): *Meaning*. Oxford: Clarendon Press.
- Scofield, J. & Behrend, D. A. (2008): «Learning words from reliable and unreliable speakers», *Cognitive Development* 23, 278-290.
- Searle, J. (1969): *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.
- Shafir, E., Simonson, I. & Tversky, A. (1993): «Reason-based choice», *Cognition* 49, 11-36.
- Shapin, S. (1994): *A Social History of Truth: Civility and Science in Seventeenth-Century England*. Chicago, IL: University of Chicago Press.
- Southgate, V., Senju, A. & Csibra, G. (2007): «Action anticipation through attribution of false belief by 2-year-olds», *Psychological Science* 18, 587-592.
- Sperber, D. (1975): *Rethinking Symbolism*. Cambridge: Cambridge University Press.
- Sperber, D. (1985): *On Anthropological Knowledge*. Cambridge University Press Cambridge.
- Sperber, D. (1996): *Explaining Culture: A Naturalistic Approach*. Oxford: Blackwell.
- Sperber, D. (1997): «Intuitive and reflective beliefs», *Mind & Language* 12, 67-83.
- Sperber, D. (2001): «An evolutionary perspective on testimony and argumentation», *Philosophical Topics* 29, 401-413.
- Sperber, D. (2010): «The Guru effect», *Review of Philosophy and Psychology* 1(4), 583-592.
- Sperber, D. & Mercier, H. (2012): «Reasoning as a social competence», in Elster, J. & Landemore, H. (eds), *Collective Wisdom*. Cambridge, Cambridge University Press, 368-392.
- Sperber, D. & Wilson, D. (1995): *Relevance: Communication and Cognition*, 2nd edn. Oxford: Blackwell.
- Sperber, D. & Wilson, D. (2005): «Pragmatics», in Jackson, F. & Smith, M. (eds), *Oxford Handbook of Contemporary Philosophy*. Oxford: Oxford University Press: 468-501.
- Stanovich, K. E. (2004): *The Robot's Rebellion*. Chicago, IL: Chicago University Press.
- Strawson, P. F. (1964): «Intention and convention in speech acts», *The Philosophical Review* 73, 439-460.
- Sunstein, C. R. (2002): «The law of group polarization», *Journal of Political Philosophy* 10, 175-195.

- Surian, L., Caldi, S. & Sperber, D. (2007): «Attribution of beliefs by 13-month-old infants», *Psychological Science* 18, 580-586.
- Talwar, V. & Lee, K. (2002): «Development of lying to conceal a transgression: children's control of expressive behavior during verbal deception», *International Journal of Behavioral Development* 26, 436-444.
- Tesser, A. (1978): «Self-generated attitude change», in Berkowitz, L. (ed.), *Advances in Experimental Social Psychology*. New York: Academic Press, 289-338.
- Thagard, P. (2002): *Coherence in Thought and Action*. Cambridge, MA: MIT Press.
- Tirole, J. (1996): «A theory of collective reputations», *The Review of Economic Studies* 63, 1-22.
- Trommershauser, J., Maloney, L. T. & Landy, M. S. (2008): «Decision making, movement planning and statistical decision theory», *Trends in Cognitive Sciences* 12, 291-297.
- Vrij, A. (2000): *Detecting Lies and Deceit: The Psychology of Lying and Implications for Professional Practice*. Chichester: John Wiley & Sons.
- Vrij, A. (2004): «Why professionals fail to catch liars and how they can improve», *Legal and Criminological Psychology* 9, 159-183.
- Walker, M. B. & Andrade, M. G. (1996): «Conformity in the Asch task as a function of age», *Journal of Social Psychology* 136, 367-372.
- Welch-Ross, M. K. (1999): «Interviewer knowledge and preschoolers' reasoning about knowledge states moderate suggestibility», *Cognitive Development* 14, 423-442.
- Willingham, D. T. (2008): «Critical thinking: why is it so hard to teach?», *Arts Education Policy Review* 109, 21-32.
- Willis, J. & Todorov, A. (2006): «First impressions: making up your mind after a 100-ms exposure to a face», *Psychological Science* 17, 592-598.
- Wilson, D. & Sperber, D. (2004): «Relevance Theory», in Horn, L. & Ward, G. (eds), *The Handbook of Pragmatics*. Oxford: Blackwell, 607-632.
- Wimmer, H. & Perner, J. (1983): «Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception», *Cognition* 13, 41-68.
- Yamagishi, T. (2001): «Trust as a form of social intelligence», in Cook, K. (ed.), *Trust in Society*. New York: Russell Sage Foundation, 121-147.
- Ybarra, O., Chan, E. & Park, D. (2001): «Young and old adults' concerns about morality and competence», *Motivation and Emotion* 25, 85-100.



# Manipulation et cognition: un modèle pragmatique

Didier MAILLAT, Université de Fribourg

## 1. Introduction

Dans son approche des phénomènes langagiers de construction de sens en contexte, la pragmatique est inévitablement entrée en contact avec les formes de discours manipulatoires et leur analyse.<sup>1</sup> Dans la plupart des cas, ce type de discours a été apparenté à une forme déficiente de la communication humaine. Les auteurs ont ainsi relevé nombre de défauts dans les énoncés manipulatoires à des niveaux différents. Le but de ce chapitre est de proposer un modèle du discours manipulatoire qui soit à même de capturer dans une théorie pragmatique la spécificité des processus qui sous-tendent la manipulation. En ce sens, ce chapitre tente de remédier au manque relevé par Cummings (2004) lorsqu'elle souligne qu'il existe une nécessité urgente à établir une théorie pragmatique systématique de l'argumentation. Ainsi, en se penchant sur les arguments réputés déficients – les arguments fallacieux – le présent chapitre apporte les premiers éléments de réponse à ce que pourrait constituer une telle théorie.<sup>2</sup>

Dans cette approche, je vais m'attacher à mettre l'accent principal de l'analyse sur la composante réceptive de la communication. Je souhaite ainsi suggérer qu'une définition adéquate de la manipulation devrait se pencher avant tout sur les phénomènes qui touchent le destinataire. La manipulation y est donc définie davantage comme un phénomène que l'on subit plus que comme quelque chose que l'on inflige. En vue de la mise en place d'un tel

1 Les idées et arguments avancés dans ce chapitre ont bénéficié des commentaires et suggestions faites par trois lecteurs anonymes, ainsi que des nombreuses discussions que j'ai eues avec Steve Oswald. Je les en remercie, même si cette dernière mouture n'engage évidemment que moi.

2 Une première version des idées présentées dans ce chapitre se trouve dans Maillat (2013).

modèle, l'effort analytique va donc être déplacé d'une approche qui présente la manipulation de façon négative comme une forme de communication déficiente dans une de ces dimensions constitutives, vers une approche qui présente la manipulation comme un effet interprétatif prévisible – dans une certaine mesure inévitable – issu d'un processus interprétatif en tout point conforme à une situation normale de communication. Dans cette entreprise, je m'attacherai à montrer que le modèle pertinentiste semble le mieux équipé pour expliciter les mécanismes qui assurent le succès du discours manipulateur. Dans la dernière section, l'analyse d'un mouvement argumentatif manipulateur bien connu – l'argument *ad populum* – est proposée afin de montrer que si un tel argument fallacieux est effectivement biaisé, il n'en est pas moins parfaitement valide d'un point de vue pragmatique et cognitif.

## 2. Quelle définition pour la manipulation?

La manipulation a souvent été présentée comme un type de communication «bancal». Ainsi, bon nombre des définitions existantes ont tenté de définir la manipulation comme déficiente à un niveau ou à un autre (voir Maillat & Oswald 2009 pour une discussion plus détaillée). La nature même de cette déficience varie. Rigotti (2005) par exemple s'est intéressé aux limites véridictionnelles des arguments manipulateurs. Les mensonges constituent à ce titre une illustration prototypique de cette forme d'usages manipulateurs. Une telle approche se heurte néanmoins à certains obstacles. On en retiendra deux en particulier: premièrement, il existe des formes d'arguments manipulateurs qui ne se caractérisent pas par des inconsistances véridictionnelles. On pense par exemple à l'utilisation d'un sentiment de peur pour renforcer un argument (comme lorsqu'on menace quelqu'un d'un *je ne ferais pas cela si j'étais vous* ou dans le cas d'un argument *ad baculum*). Deuxièmement, même dans les cas où les conditions de vérité sont ouvertement violées, par exemple avec un argument *ad populum* du genre

- (1) Tout le monde pense que Hollande sera le prochain président,



il semble que la véritable intention manipulative ne réside pas dans l'entorse à la vérité qui y est faite (à savoir que (1) n'est *pas* une pensée partagée par tout le monde). Dans notre exemple, la force de l'usage manipulative et son effet sont ainsi préservés même si l'on réduit la portée référentielle du quantificateur universel:

- (2) Tout le monde à qui j'ai parlé (Toutes les personnes à qui j'ai parlé) pense(nt) que Hollande sera le prochain président.

En effet, sur un plan rationnel et selon les théories de l'argumentation, (2) reste un énoncé manipulative (un argument *ad populum*) qui présente le nombre d'individus qui partagent une pensée rapportée comme un argument en faveur de la validité intrinsèque de cette pensée, ignorant en cela le fait que tout le monde, respectivement tout-le-monde-à-qui-j'ai-parlé, peut être dans l'erreur d'un point de vue rationnel.

D'autres comptes rendus, comme celui de van Dijk (2006), proviennent de la tradition d'analyse critique du discours, où les analystes se penchent sur les relations de pouvoir qui sont négociées et imposées au fil du discours. Dans ce paradigme, le discours manipulative est conçu prioritairement comme un effort de la part de la locutrice visant à maintenir et à garantir sa propre position sociale dominante en feignant de s'intéresser aux intérêts du destinataire, alors même que seuls les intérêts de la locutrice sont en jeu (voir également Saussure 2005). Une telle approche paraît, elle aussi, trop restrictive dans la mesure où elle exclut des contextes énonciatifs dans lesquels une locutrice souhaite manipuler un destinataire pour le bien de ce dernier.

En pragmatique, on trouve des chercheurs qui ont essayé de circonscrire la spécificité du discours manipulative en utilisant les outils traditionnels de la discipline. Parret (1978) par exemple évalue le potentiel d'une approche par les actes de langage, au sein de laquelle la déficience des formes manipulatives est envisagée sur le plan de la violation des conditions de félicité. Toutefois, que la manipulation soit présentée comme un phénomène perlocutoire ou qu'elle soit présentée comme une force illocutoire, l'idée même d'un acte de langage *manipulation* (parallèlement à celui de *promesse*, ou de *menace*, par exemple) se heurte à un problème considérable puisqu'il s'agit d'un type d'acte de langage dont la force et la portée perlocutoire seraient invariablement anéanties au moment précis où le type d'acte – la *manipulation* – serait identifié. En d'autres termes, un acte de *manipulation* constituerait une

aberration en ce sens que pour viser à une quelconque félicité, il devrait passer inaperçu (voir la discussion chez Parret 1978).

Mentionnons enfin un effort dont les hypothèses de base sont apparentées au modèle proposé dans le présent chapitre. Certains chercheurs ont ainsi mis l'accent sur le décalage existant entre la visée communicative d'une locutrice manipulatrice et l'intention qui lui est attribuée par le destinataire (Attardo 1997, Rigotti 2005), en insistant sur le décalage intentionnel (intention sous-jacente vs intention affichée) qui sous-tend la production d'un énoncé manipulateur. Toutefois, ces approches intéressantes rencontrent elles aussi un obstacle puisqu'elles définissent la manipulation comme quelque chose que la locutrice fait, alors même qu'on peut montrer que la manipulation peut s'exercer sans décalage intentionnel (comme dans le cas où un message de propagande est répété par des foules endoctrinées).

Par exemple, on peut penser à une situation où un ensemble d'arguments manipulateurs peuvent produire le même effet manipulateur chez un destinataire indépendamment du locuteur qui les relaie au destinataire. Les arguments fallacieux racistes peuvent ainsi s'appuyer sur des preuves de nature anecdotique pour justifier une législation raciste qui impose une hiérarchie sociale fondée sur les origines raciales. On peut légitimement supposer que les initiateurs d'une telle stratégie de propagande essaient de convaincre leur public cible en le manipulant. L'intention manipulatrice sous-jacente pourrait être décrite comme visant à mettre en place un sentiment de supériorité et d'animosité à l'encontre d'une minorité sociale. Mais si l'argument peut être fallacieux dans l'esprit des initiateurs, il est également très probable que parmi les nombreuses personnes qui, une fois convaincues, relayeront l'argument, ce dernier n'est pas perçu comme fallacieux. Dès lors, quoique l'argument fallacieux continue à se propager, l'intention manipulatrice disparaît chez les locuteurs. Dans un tel contexte, il serait impossible d'analyser la manipulation – qui a pourtant lieu – en termes d'intentions cachées des locuteurs.

Par exemple, en 1933, les leaders nazis savaient que si une fausse assertion est répétée suffisamment souvent, elle deviendrait une «vérité». Quoiqu'il soit clair qu'une telle stratégie de manipulation (voir la discussion plus loin de l'argument fallacieux *ad populum*) reflète le type d'intention qui motivait les dirigeants nazis lorsqu'ils planifièrent la propagande autour de l'incendie du

Reichstag,<sup>3</sup> une telle intention cachée était certainement absente dans l'esprit des masses qui ont contribué à la propagation rapide de ces arguments manipulateurs. Il n'en reste pas moins que toute personne ayant été exposée à ce discours peut être considérée comme ayant été la victime d'une manipulation (indépendamment de l'intention du locuteur). Avec un tel exemple, il apparaît donc qu'un énoncé manipulateur peut être relayé et produire les mêmes effets même lorsque l'intention manipulative disparaît. L'efficacité même des techniques de propagande réside du reste sur cette propriété des arguments manipulateurs.

Au terme de ce rapide survol des modèles proposés,<sup>4</sup> un des obstacles récurrents rencontrés par ces derniers naît de leur insistance à définir la manipulation avant tout du point de vue de la locutrice, ce qui les conduit à des définitions s'appuyant sur des notions de force, mouvement, ou intention cachée qui compliquent, voire empêchent, une analyse explicite du phénomène. Dans ce qui suit, je propose de contourner cet obstacle afin d'arriver à une analyse plus riche des usages langagiers manipulateurs en inversant la perspective et en considérant ces phénomènes du point de vue de leur destinataire et des phénomènes interprétatifs qu'ils déclenchent chez ce dernier.

### 3. Pragmatique de la manipulation

Une approche pragmatique des énoncés manipulateurs peut être proposée comme alternative. En particulier, l'option retenue ici envisage la manipulation comme un effet portant sur les processus interprétatifs du destinataire plutôt que comme un usage trompeur de la part de la locutrice. En modifiant ainsi la perspective, on rend les usages manipulateurs plus accessibles d'un point de vue analytique, par rapport à une approche qui se focalise sur une intention masquée de la locutrice.

3 Les activistes communistes avaient été désignés (à tort) comme les seuls coupables.

4 Le lecteur intéressé consultera notamment Saussure & Schultz (2005) et Maillat & Oswald (2009) pour une discussion plus approfondie de la littérature.

Un tel changement de perspective entraîne notamment une modification non-triviale du type de définition qui peut être mis en avant pour le discours manipulateur. La différence la plus remarquable concerne la nature même de la communication manipulateur. En effet, du côté du destinataire, un énoncé manipulateur est traité comme tout autre énoncé. En d'autres termes, les processus interprétatifs déclenchés par un énoncé manipulateur sont très exactement ceux décrits pour les situations «normales» de communication. En fait, comme je vais le montrer plus loin, c'est précisément la nature non-marquée de l'interprétation d'un énoncé manipulateur qui en assure le succès. Toutefois, il serait faux – et contre-productif – de conclure que l'interprétation du discours manipulateur ne se distingue pas des cas de discours normal. L'idée défendue dans ce chapitre consiste à définir la manipulation comme un résultat sous-optimal du processus normal d'interprétation. C'est-à-dire qu'il s'agit d'une situation où le mécanisme d'interprétation normal se fourvoie. La différence fondamentale qui existe entre ce type de processus interprétatif erroné et un cas de fausse interprétation accidentelle réside dans le fait que le résultat sous-optimal généré par la manipulation est considéré comme étant obtenu par l'application de contraintes cognitives appliquées sur les mécanismes qui régissent l'interprétation.<sup>5</sup>

Afin de pouvoir décrire ce phénomène, un cadre d'analyse pragmatique semble particulièrement approprié. La théorie de la pertinence (Blakemore 2002, Carston 2002, Sperber & Wilson 1995, Wilson & Sperber 2012) apparaît en effet comme le modèle le plus adéquat pour le type d'analyse requis. Le premier argument qui parle en faveur d'un modèle pertinentiste relève de son traitement du contexte. Comme Sperber & Wilson (1995) le soulignent, un des points de divergence avec les autres modèles pragmatiques repose dans la façon d'intégrer le contexte au processus interprétatif. Ils écrivent que «[i]t is not that first context is determined, and then relevance is assessed. [...] [I]t is relevance which is treated as given, and context which is treated as a variable» (1995: 141-142). Ils expliquent ainsi que le contexte est la résultante du processus interprétatif qui conduit à l'interprétation de l'énoncé.

5 Une telle distinction est nécessaire afin de ne pas traiter comme un cas de manipulation une situation où un énoncé comme *Max regarde la glace* est interprété comme signifiant que Max regarde un dessert, alors que c'est un miroir que ce dernier observe.

Une telle approche est intéressante pour un traitement des usages manipulateurs dans le sens où elle recentre le processus interprétatif autour des mécanismes de sélection de contexte. Par là, la théorie de la pertinence prédit également que l'interprétation va dépendre de ces mécanismes de sélection et qu'une stratégie visant à contrôler les mécanismes de sélection de contexte équivaldrait à contrôler l'interprétation obtenue par le destinataire. En d'autres termes, dans un cadre théorique pertinentiste, la manipulation peut être analysée comme une tentative de contrôle du processus de sélection de contexte du destinataire. C'est l'option analytique retenue dans ce chapitre.

La pertinence n'est donc pas le but visé par le processus interprétatif, mais le principe directeur qui régit la sélection de contexte (1995: 141-142): une interprétation optimale minimise les efforts cognitifs que le destinataire doit investir dans la construction de l'ensemble d'informations contextuelles, et maximise les effets cognitifs générés par cet ensemble contextuel. On remarquera en particulier, comme le soulignent Sperber & Wilson (1995), que la construction de l'ensemble d'informations contextuelles est un processus incrémentiel par lequel le destinataire accède à des éléments informatifs dans son environnement cognitif en suivant un ordre d'accessibilité. Il s'ensuit que toutes les informations qui sont manifestes pour un individu au moment où il interprète un énoncé ne sont pas toutes aussi accessibles les unes que les autres; certaines informations sont plus manifestes que d'autres (1995: 39-40). Comme la recherche de pertinence optimale ne sélectionne qu'une fraction de toutes les informations manifestes au destinataire, il s'ensuit que même dans l'hypothèse improbable où deux destinataires partagent un environnement cognitif identique sur le plan des informations qui leur sont manifestes, il existera vraisemblablement une différence quant au degré d'accessibilité de ces informations dans leur environnement cognitif respectif. En conséquence, un même énoncé *U* peut conduire à une interprétation différente pour ces deux individus simplement parce que d'autres informations seront sélectionnées en premier en fonction de leur niveau d'accessibilité.

Sperber & Wilson (1995) tentent de saisir l'effet de la relation d'ordre qui structure l'environnement cognitif au moyen de différentes notions parfois proches les unes des autres. Ils mentionnent ainsi des degrés de manifesteté (*manifestness*), d'accessibilité (1995: 138), et également de force d'une information (1995: 75sq, 143, 151). Plus que le détail de l'enchaînement des

différents niveaux contextuels ce sont les propriétés mêmes de la structure ordonnée de l'environnement cognitif qui nous importent ici. En particulier, en vue d'une meilleure compréhension des usages manipulateurs, on relèvera que

*The organisation of the individual's encyclopaedic memory, and the mental activity in which he is engaged, limit the class of potential contexts from which an actual context can be chosen at any given time. [...] [N]ot all chunks of encyclopaedic information are equally accessible at any given time.* (Sperber & Wilson 1995: 138)

Du fait que le processus d'optimisation au cœur du principe de pertinence n'est pas exhaustif, des différences dans la relation d'ordre qui structure l'environnement cognitif conduisent à une sélection différente de l'ensemble contextuel,<sup>6</sup> et donc à une interprétation différente. La manipulation, en ce sens, est un usage discursif qui consiste à contrôler la relation d'ordre qui structure les informations dans l'environnement cognitif du destinataire en fonction de leur degré d'accessibilité.

Le deuxième argument en faveur d'un modèle théorique pertinentiste est lié à la spécification explicite dans le modèle d'un niveau-seuil où les processus d'enrichissement pragmatiques s'interrompent pour livrer une interprétation perçue comme optimale. En particulier, la théorie de la pertinence insiste sur le fait que la pertinence est calculée comme un ratio optimal entre efforts et effets cognitifs, mais la notion d'optimalité utilisée n'est pas considérée comme absolue. Elle est déterminée relativement à un environnement cognitif et à la disponibilité de ressources cognitives. En ce sens, le premier principe de pertinence est décrit comme une *tendance* vers la pertinence optimale (voir la discussion dans Sperber & Wilson 1995: 262). Mais Sperber & Wilson s'empressent de préciser que «There may be many shortcomings, many cognitive sub-mechanisms that fail to deliver enough effect for the effort they require, many occasions when the system's resources are poorly allocated» (*ibid.*).

En d'autres termes, le système pragmatique présenté dans la théorie de la pertinence est un système susceptible de se tromper. Non seulement l'erreur est possible, mais la théorie fait même un certain nombre de prédictions

6 Sperber & Wilson précisent que «A phenomenon may make manifest a very large number of assumptions. However, this is not to say that the individual will actually construct any, let alone all, of these assumptions» (1995: 151).

quant au type d'erreurs qui sont susceptibles d'affecter le rendement du système et de conduire à une interprétation sous-optimale. L'argument principal de ce chapitre pose que le discours manipulateur exploite les faiblesses inhérentes au processus interprétatif pour faire en sorte qu'une interprétation sous-optimale soit dérivée, c'est-à-dire que la manipulatrice s'assure qu'une des erreurs types prédites par le système se produise.<sup>7</sup> Dans ce contexte, l'approche proposée des usages manipulateurs considère ces derniers comme la conséquence inévitable de la façon dont fonctionne notre système pragmatique.

Plus précisément, cette analyse suggère qu'une manipulatrice va chercher à atteindre son but discursif en réordonnant l'environnement cognitif du destinataire de manière à garantir qu'un énoncé *E* sera interprété dans le cadre d'un sous-ensemble d'éléments contextuels donné, et en ignorant la présence d'informations contradictoires au sein de ce même environnement cognitif. Il ressort qu'en ce qui concerne la manipulation, d'un point de vue interprétatif, la composante centrale d'un argument manipulateur est une contrainte appliquée au processus de sélection des informations contextuelles en vue de l'interprétation d'un énoncé-cible *E*. La manipulation est donc ré-analysée comme une instance de *Contrainte de Sélection Contextuelle* (ou CSC).

La Manipulation comme Contrainte de Sélection Contextuelle – une définition:

La communication manipulateur est un processus double par lequel une contrainte qui limite la sélection de contexte est combinée avec un énoncé-cible *E* de manière à forcer l'interprétation de ce dernier dans un sous-ensemble d'informations contextuelles restreint et d'assurer que l'interprétation est atteinte avant que le destinataire n'accède à un autre sous-ensemble contextuel connu et contradictoire avec *E*.

Avec ce type d'analyse, deux stratégies de contrainte de sélection contextuelle sont envisageables pour un usage manipulateur. Selon Sperber & Wilson (1995: 140-141), l'environnement cognitif du destinataire est une structure ordonnée soumise à une relation partielle d'inclusion entre les sous-ensembles qui la composent. Au sein de cette structure ordonnée, les informations dérivées de l'interprétation de l'énoncé précédent constituent un sous-ensemble contextuel immédiatement accessible. A ce niveau initial

7 Conformément à l'usage retenu dans de nombreux articles et ouvrages pertinents, l'interaction verbale retenue met en contact une locutrice et un destinataire.

s'ajoute un sous-ensemble contextuel étendu qui inclut «the encyclopaedic entries (or possibly smaller chunks of encyclopaedic information, taken from these entries) of concepts already present either in the context or in the assumption being processed». Enfin, un troisième niveau est identifié par Sperber & Wilson qui englobe les informations sur l'environnement immédiatement observable. Mais ce qu'il importe de relever en particulier dans cette présentation pertinentiste de la structure de l'environnement cognitif réside dans le fait que cette relation formelle, existant entre différents sous-ensembles,

*has a psychological counterpart: order of inclusion corresponds to order of accessibility [...]; contexts which include only the initial context as a sub-part [...] are therefore the most accessible contexts; contexts which include the initial context and a one-step extension as sub-parts [...] are therefore the next most accessible contexts, and so on. (1995: 142)*

En revenant à la notion de manipulation, l'approche retenue ici pose que les usages manipulateurs s'assurent qu'un contexte  $C'$  qui serait pertinent à l'interprétation d'un énoncé  $E$  ne soit pas accessible pour le destinataire dans la mesure où il conduirait à une contradiction avec  $E$  et au-delà à l'élimination de  $E$  de l'environnement cognitif du destinataire ( $EC^D$ ). Au lieu de cela, la manipulation vise à rendre un contexte  $C$ , au sein duquel  $E$  est renforcé et donne lieu à des inférences contextuelles, beaucoup plus accessible, et par là-même à bloquer l'accès à  $C'$ , puisque «[t]he less accessible a context, the greater the effort involved in accessing it, and conversely» (1995: 142). Ainsi, la manipulation exploite-t-elle le fait que la pertinence est plus susceptible d'être atteinte dans des sous-ensembles contextuels hautement accessibles.<sup>8</sup>

Le lecteur aura déjà remarqué qu'une analyse de la manipulation qui se concentre sur les effets interprétatifs produits sur le destinataire et qui ne se concentre pas sur la stratégie masquée de la locutrice mentionnée dans nombre de modèles existants se heurte à une difficulté quand il s'agit de distinguer la manipulation d'un processus interprétatif normal, ou encore, pour utiliser une terminologie propre à la théorie de l'argumentation (van Eemeren & Grootendorst 2004), de distinguer manipulation et persuasion.

8 Maillat (2006) et Maillat & Oswald (2009) montrent comment une stratégie manipulateuse peut utiliser la sur-accessibilité d'un contexte (par exemple 9/11 au lendemain des attaques contre le World Trade Center) afin de s'assurer que le contexte  $C'$  soit simplement ignoré.



Dans le modèle proposé, la persuasion est une tentative discursive de renforcer un énoncé  $E$  dans  $EC^D$  en s'assurant que  $E$  est consistant et, idéalement, renforcé dans tous les sous-ensembles pertinents susceptibles d'être accessibles au sein de  $EC^D$ .

$\neg \exists C', C' \subset EC^D$  et  $C' \neq C$ , tel que  $E$  est pertinent dans  $C'$  et  $E$  est inconsistant et éliminé dans  $C'$

En d'autres termes, la persuasion vise à ce qu'il n'existe aucun sous-ensemble contextuel  $C'$  dans lequel  $E$  est à la fois pertinent et inconsistant. Lorsqu'un destinataire est persuadé par  $E$ , l'hypothèse contextuelle correspondante est renforcée dans tous les sous-ensembles de  $EC^D$  dans lesquels  $E$  est pertinent.

La manipulation, quant à elle, est une tentative discursive de garantir que tout autre sous-ensemble  $C'$  connu et pertinent de  $EC^D$ , lequel conduirait à une élimination de  $E$ , ne soit pas accessible – ou du moins qu'il soit moins accessible que  $C$ .

$\forall C', C' \subset EC^D$  et  $C' \neq C$ , tel que  $E$  est pertinent dans  $C'$  et  $E$  est inconsistant et éliminé dans  $C'$ , accessibilité ( $C$ ) > accessibilité ( $C'$ ) dans  $EC^D$

Une analogie peut ici nous aider à expliciter ce contraste. Les actions de persuasion et de manipulation peuvent être rapprochées de deux actions qui s'offrent à une personne lorsqu'elle souhaite faire traverser une frontière à certains documents. La persuasion correspond à une tentative de la part de la locutrice de dédouaner une information dans  $EC^D$  en obtenant le bon visa pour cette dernière, alors que la manipulation est une tentative de faire passer de façon illégale cette même information dans  $EC^D$  («getting past [the hearer's] defences» selon Sperber et al. 2010).

Sur la base de ce modèle, deux types de stratégies de Contrainte de Sélection Contextuelle sont susceptibles de permettre à la manipulatrice de déclencher l'effet recherché chez le destinataire. Elle peut soit choisir de renforcer l'accessibilité du sous-ensemble contextuel  $C$  dans lequel  $E$  est renforcé, ou au contraire d'affaiblir l'accessibilité du sous-ensemble contextuel  $C'$  dans lequel  $E$  est inconsistant et éliminé.<sup>9</sup> La Figure 1 illustre ces deux stratégies.

9 Les principes qui gouvernent la pertinence nous laissent envisager deux types de manœuvres pour affaiblir ou renforcer un contexte: un contexte sera affaibli dans la

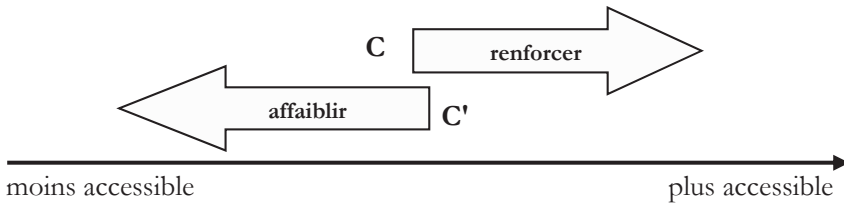


Figure 1 Deux formes de Contraintes de Sélection Contextuelle: affaiblir et renforcer le degré d'accessibilité

Après avoir présenté le cadre théorique d'une analyse pragmatique du discours manipulateur, un exemple prototypique d'argument fallacieux manipulateur est analysé afin d'évaluer le potentiel explicatif du modèle et de souligner sa convergence avec d'autres aspects de la théorie de la pertinence.

#### 4. Argument fallacieux pas si fallacieux: le cas de l'*ad populum*

Maillat & Oswald (2009, 2011) explorent l'interface entre théorie de l'argumentation et psychologie cognitive. Ils posent notamment que la CSC permet à l'analyste d'expliquer les arguments manipulateurs et en particulier leurs effets sur le destinataire avec une plus grande finesse et de façon plus explicite.

Il y a plusieurs façons de mettre en œuvre l'une ou l'autre des stratégies de manipulation identifiées dans la section précédente. Cependant, toutes vont faire en sorte de modifier la structure interne de l'environnement cognitif. En ce sens, la manipulation se positionne clairement comme un phénomène cognitif, ce qui ne constitue pas une surprise si on en accepte les fondements pragmatiques dans une théorie pragmatique de la pertinence qui

mesure où plus d'effort cognitif sera nécessaire pour accéder aux hypothèses contextuelles qui le constituent (et inversement pour renforcer un contexte) ; un contexte sera affaibli dans la mesure où il permettra de générer moins d'effets cognitifs lors de l'interprétation de *E* (et inversement pour renforcer un contexte).

postule sa propre nature et validité cognitives. Comme il a été relevé dans une citation précédente, les mécanismes cognitifs qui activent – entre autres – les processus interprétatifs sont susceptibles d’erreurs.

Une bonne illustration de ces limitations qui affectent les processus cognitifs se trouve dans la littérature de psychologie cognitive, plus précisément autour de la notion d’illusion cognitive (voir Pohl 2004 pour un survol détaillé). Allott & Rubio Fernandez (2002) se sont également intéressés aux implications pragmatiques de telles limitations cognitives en liaison avec ce que l’on nomme couramment l’illusion de Moïse (Reder & Kusbit 1991). Pohl (2004: 2-4) définit les illusions cognitives comme un phénomène cognitif qui i) s’écarte d’une norme «correcte», ii) s’en écarte de façon systématique, iii) échappe à un contrôle conscient, et de ce fait, iv) est difficile à éviter. Il n’est dès lors pas étonnant de constater que nombre de stratégies manipulatoires exploitent précisément ces illusions cognitives de manière à induire le système cognitif en erreur (Maillat & Oswald 2009 identifient plusieurs de ces illusions manipulatoires).

D’après van Emmeren & Houtlosser (2008: 140), l’argument fallacieux *ad populum* peut être défini comme l’argument fallacieux qui consiste «of regarding something acceptable because it is considered acceptable by a great many people». En d’autres termes, l’argument *ad populum* se fonde sur le fait que la répétition de *E* par plusieurs sources distinctes tend à renforcer *E*. Il n’est pas nécessaire de chercher très longtemps pour trouver un exemple d’usage manipulatoire de ce type, puisque l’histoire de la propagande et de la publicité en abonde. Les campagnes de pub de type «burst advertising» par exemple se fondent sur la présence envahissante et le martelage répété d’un slogan pour renforcer son impact sur le public cible. De même, le régime nazi s’est assuré que la population allemande soit exposée de manière répétée aux credos du parti (par exemple concernant la supposée infériorité biologique de certaines minorités ethniques, religieuses, ou politiques) de façon à renforcer leur accessibilité dans ECP.

Dans une perspective CSC, l’analyse des processus cognitifs sous-jacents à un argument fallacieux *ad populum* est relativement claire. La répétition de *E* garantit un plus haut degré d’accessibilité. *E* en sort renforcé dans ECP à chaque nouvelle occurrence. Dans le modèle de CSC, ceci correspond à une contrainte de renforcement de *C* par répétition qui assure que les sous-

ensembles contextuels inconsistants (qui conduiraient à une élimination de *E*) ne sont pas accédés.

Il existe par ailleurs un faisceau considérable de preuves empiriques dans les champs de la psycholinguistique et de la psychologie cognitive qui pointe vers une confirmation de l'effet cognitif (sur ECP) noté plus haut. On mentionnera ici deux directions explorées par les chercheurs qui ont investigué les mécanismes cognitifs qui sous-tendent les phénomènes de renforcement par répétition. Ces deux effets appartiennent à la catégorie des illusions cognitives.

La psychologie cognitive a ainsi identifié «l'effet de simple exposition» (*mere-exposure effect*) qui est défini comme un «increased liking for a stimulus that follows repeated, unreinforced exposure to that stimulus» (Bornstein & Craver-Lemley 2004: 231). L'effet de simple exposition a été confirmé pour des stimuli très différents (visuels, acoustiques, linguistiques, entre autres). Si la notion de jugement positif ne correspond pas directement à celle d'accessibilité, j'émetts l'hypothèse que l'augmentation observée de connotation positive associée avec une certaine information correspond à une plus grande accessibilité. Cet effet est le résultat du fait que le destinataire est spontanément moins vigilant – au sens de Sperber et al. 2010 – à l'égard d'hypothèses contextuelles qu'il juge positivement (voir Zajonc (1968) pour une discussion du «mere-exposure effect» au niveau lexical).

Un second effet cognitif important, lui aussi lié à un renforcement par répétition, est appelé le «validity effect». Contrairement au «mere-exposure effect», ce second type d'illusion cognitive est fermement ancré dans le langage. Hackett Renner (2004:201) décrit son impact sur l'évaluation par le destinataire d'une certaine information transmise verbalement: «*[I]f information has been heard previously, people are likely to ascribe more truth or validity to it than if they are hearing it for the first time [...] regardless of whether the information was originally believed to be true or false.*»

Ce qui revient à dire que plus une information est répétée, plus elle est perçue comme valide. En conséquence, la simple répétition de *E* peut transformer un ECP dans lequel *E* était initialement éliminé en un ECP consistant avec *E*. Ce phénomène constitue la contrepartie cognitive de l'argument fallacieux *ad populum*. Ces effets fournissent une explication cognitive à la stratégie de CSC de renforcement par répétition. Il est intéressant de noter que les résultats expérimentaux obtenus en

psycholinguistique nous montrent que si l'argument *ad populum* est un argument déficient d'un point de vue rationnel (rien ne permet de garantir la validité d'un argument par le nombre d'individus qui y souscrivent), il n'en constitue pas moins un argument parfaitement valide d'un point de vue cognitif puisque la répétition de *E* garantit effectivement que l'hypothèse contextuelle dérivée de *E* sera renforcé dans  $EC^D$ . En conséquence, si tout le monde dit *E*, l'hypothèse contextuelle dérivée de *E* devient effectivement plus accessible dans  $EC^D$  et sera donc plus susceptible d'être intégrée dans un processus interprétatif.

Toutefois, si cette analyse explique pourquoi et comment la propagande ou la publicité s'appuient sur une exposition répétée à *E* afin de renforcer la force de *E* dans  $EC^D$ , elle ne parvient pas encore à expliquer pourquoi un argument fallacieux *ad populum* qui manipule le destinataire pour qu'il accepte *E* en se contentant d'affirmer que *tout le monde dit E* devrait fonctionner. Pourtant un argument comme *tout le monde dit E donc E doit être vrai/est un argument valide* illustre une des formes les plus communes et fréquentes d'argument fallacieux utilisé dans la vie de tous les jours.

De manière à comprendre ce qui se passe dans ce dernier cas, il faut commencer par relever que *tout le monde dit E donc E doit être vrai* résume ce qui se passe effectivement sur un plan cognitif si *E* est répété par tout le monde dans l'environnement du destinataire. Comme nous l'avons vu plus haut, la répétition de *E* renforce *E* dans  $EC^D$ . Dans la mesure où l'antécédent de l'énoncé fallacieux décrit effectivement un état du monde (il est exact que tout le monde dit *E*), le conséquent suit cognitivement du fait des effets de «mere-exposure» et de «validity» (*E* sera renforcé dans  $EC^D$ ). En ce sens, *tout le monde dit E donc E doit être vrai* fonctionne et peut être considéré comme le raccourci d'un renforcement par répétition. Un tel mouvement argumentatif se fonde donc sur un processus cognitif de renforcement d'une hypothèse contextuelle par simple allusion à ce processus. L'*ad populum* ne fait ainsi que mentionner la répétition sans l'utiliser, mais l'effet cognitif dans  $EC^D$  est le même.

A ce titre il est intéressant de relever que la psychologie cognitive s'est également penchée sur l'impact de la répétition en mention et non en usage. En d'autres termes, les chercheurs se sont attachés à établir si le simple fait de mentionner qu'un stimulus a été répété (répétition en mention) suffit à déclencher un effet de validité même dans le cas où le sujet n'a pas été exposé

à plusieurs occurrences du stimulus en question (pas de répétition en usage). Le collectif Bacon (1979: 247) a ainsi montré qu'un *jugement de reconnaissance* d'un stimulus donné – plutôt que sa répétition effective – suffit à produire l'effet de validité: «Unambiguously, then, differences in rated truth are more sensitive to recognition judgments than to actual repetition of the statements being rated.» Ozubko & Fugelsang (2011: 273) ont également réussi à reproduire l'effet en question en induisant les sujets en erreur sur la présence dans leur mémoire du stimulus visé. On pourrait ainsi expliquer notre exemple en disant que l'argument *ad populum* se sert de l'effet cognitif de la répétition en mention.

La théorie de la pertinence fournit encore une fois un éclairage intéressant en lien avec ces dernières observations. La TP s'est en effet intéressée aux marqueurs linguistiques qui permettent de contrôler les processus inférentiels qui guident l'interprétation. Dans son travail sur la pragmatique des connecteurs discursifs, Blakemore (2002:89) remarque que les langues ont développé des «coded means for constraining the inferential tasks involved in utterance interpretation». Ces items lexicaux sont appelés des *expressions procédurales*. Dans ce contexte, je propose d'envisager qu'un énoncé du type *tout le monde dit E* fonctionne comme une expression procédurale en ce sens qu'elle contraint le système inférentiel vers un renforcement-par-répétition, et parvient ainsi à manipuler l'interprétation de *E* (renforcement de *E* dans  $EC^D$ ). En ce sens, *tout le monde dit E* ouvre un raccourci cognitif puisque les expressions procédurales «ensure the recovery of the intended cognitive effects for a minimum cost in processing» (Blakemore 2002: 130).

## 5. Conclusion

Dans ce modèle, les effets manipulatoires sont obtenus en contraignant le processus de sélection contextuelle. Dans l'étude de cas, nous avons vu que pour les *ad populum* la contrainte appliquée sur  $EC^D$  peut prendre deux formes. La manipulatrice peut adopter une stratégie en force et décider de contraindre l'interprétation en augmentant le degré d'accessibilité de *C* par la répétition; ou elle peut avoir recours à une expression procédurale (*tout le*

*monde dit E donc E*) pour contraindre le processus inférentiel et le guider au travers d'un raccourci.

L'existence et l'intérêt de tels raccourcis cognitifs sont prédits par une théorie pragmatique comme la théorie de la pertinence qui se fonde sur des principes d'optimisation cognitive. Dans un cadre théorique pertinentiste, l'interprétation tend à l'efficacité maximale. D'un point de vue évolutionniste, on s'attend à ce que ce type d'heuristique donne lieu à des raccourcis de traitement cognitif. La manipulation est donc une stratégie qui abuse de cette attraction naturelle – inévitable – pour les raccourcis. Souvent les manipulateurs vont ainsi réussir à tromper les destinataires parce que comme le rappellent Sperber *et al* (1995: 90): «[...] *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.*»

Dans un récent développement de ces idées, Sperber *et al.* (2010) font l'hypothèse que le système cognitif a développé évolutivement un mécanisme parallèle de vigilance épistémique de manière à contrebalancer les effets de cette propension à l'optimisation. Le rôle de la vigilance épistémique est d'assurer que les mécanismes cognitifs qui régissent l'interprétation ne sont pas piégés trop souvent. En ce sens les mécanismes de la vigilance épistémique sont appelés à être intimement liés avec les usages manipulateurs du langage, puisque les deux fonctionnent dans une direction exactement inverse.

La littérature existante sur les arguments réputés fallacieux (voir van Eemeren & Grootendorst 2004) fournit de nombreux exemples de cette relation antagoniste. Sperber *et al.* (2010) évoquent deux orientations principales que peut prendre la vigilance épistémique: i) la vigilance à l'égard du contenu, et ii) la vigilance vis-à-vis de la source de l'énoncé. Comme il est attendu, on trouve des usages manipulateurs qui tentent effectivement de tromper la vigilance épistémique sur la nature de la source. Par exemple, une expression du type *les experts disent U* – une instance d'argument réputé fallacieux dans les théories de l'argumentation, dit *ad verecundiam* – peut être analysée comme une expression procédurale qui contraint l'interprétation de *E* en diminuant l'accessibilité de tout sous-ensemble contextuel *C'* dans lequel *E* est éliminé. Similairement, on trouve des expressions procédurales qui flouent la vigilance épistémique sur le contenu d'un énoncé. Les syllogismes par exemple sont des structures inférentielles sûres et le contenu

des conclusions qu'ils induisent est valide. Il n'est dès lors pas étonnant que les manipulateurs sachent qu'un argument qui ressemble à un syllogisme a de bonnes chances d'être accepté par le filtre de vigilance épistémique. Pour cette raison, la manipulation fait un usage immodéré d'arguments fallacieux déguisés en syllogismes, comme l'affirmation du conséquent par exemple.

- (6) Vous savez ce qu'ils disent. «Si on voit Lou quelque part, Max n'est pas bien loin.» – Et je viens de voir Max... donc Lou doit être ici.

Ici, la structure quasi-syllogistique exploite la contrainte inférentielle générée par les véritables syllogismes qui fonctionnent comme des expressions procédurales de forme *si P alors Q; P; donc Q*.<sup>10</sup>

Au travers de ces exemples, je n'ai pu donner que quelques indications du potentiel analytique de ce modèle, mais j'espère avoir établi que ce type d'approche pragmatique des usages manipulateurs du langage ouvre une perspective nouvelle et riche dans des débats qui occupent les linguistes depuis Aristote.

## Bibliographie

- Allott, N. & Rubio Fernandez, P. (2002): «This paper fills a much-needed gap», in Afuta, P., El Ghali, A. & Toussnel, F. (éds.), *Actes de l'Atelier des doctorants en linguistique*, Université Paris 7, 97-102.
- Attardo, S. (1997): «Locutionary and Perlocutionary Cooperation: The Perlocutionary Cooperative Principle», *Journal of Pragmatics* 27, 753-779.
- Bacon, F. T. (1979): «Credibility of repeated statements: Memory for trivia», *Journal of Experimental Psychology: Human Learning and Memory*, 5(3), 241-252.
- Blakemore, D. (2002): *Relevance and linguistic meaning: the semantics and pragmatics of discourse markers*. Cambridge University Press, Cambridge.

- 10 Traditionnellement, ce sont les marqueurs discursifs qui ont été traités comme expressions procédurales (par exemple *donc* dans mon exemple). Toutefois, dans la mesure où un syllogisme peut se construire sans connecteur et que les participants sont néanmoins capables de l'identifier (*Tous les hommes sont mortels. Socrates est un homme. Socrates est mortel*), c'est l'identification d'une structure syllogistique ou quasi-syllogistique qui produit le raccourci cognitif dont il est question ici. Or cette structure syllogistique est indépendante de la présence de connecteurs comme *donc* dans l'énoncé.



- Bornstein, R. F. & Craver-Lemley, C. (2004): «Mere exposure effect», in Pohl, R. F. (éd.), *Cognitive Illusions: A Handbook on Fallacies and Biases in Thinking, Judgement and Memory*. Psychology Press, Hove, 215–234.
- Carston, R. (2002): *Thoughts and Utterances: The Pragmatics of Explicit Communication*. Blackwell, Oxford.
- Cummings, L. (2004): *Pragmatics: a multidisciplinary perspective*. L. Erlbaum Associates, Mahwah, N.J.
- Eemeren, F. H. van & Grootendorst, R. (2004): *A Systematic Theory of Argumentation: the pragma-dialectical approach*. Cambridge University Press, Cambridge.
- Eemeren, F. H. van & Houtlosser, P. (2008): «Rhetoric in a Dialectical Framework: Fallacies as Derailments of Strategic Manoeuvring», in Weigand, E. (éd), *Dialogue and Rhetoric*. John Benjamins, Amsterdam, 133-152.
- Grice, H.P. (1989): *Studies in the way of words*. Harvard University Press, Harvard.
- Hackett Renner, C. (2004): «Validity Effect», in Pohl, R. F. (éd.), *Cognitive Illusions: a handbook on fallacies and biases in thinking, judgment and memory*. Psychology Press, Hove, 201-213.
- Levinson, S. (2000): *Presumptive Meanings: The theory of generalized conversational implicature*. MIT Press, Cambridge, MA / London.
- Maillat, D. (2013): «Constraining context selection: on the pragmatic inevitability of manipulation», *Journal of Pragmatics* 59B, 190-199.
- Maillat, D. & Oswald, S. (2009): «Defining manipulative discourse: the pragmatics of cognitive illusions», *International Review of Pragmatics* 1(2), 348-370.
- (2011): «Constraining context: a pragmatic account of cognitive manipulation», in Hart, C. (éd.), *Critical discourse studies in context and cognition*. John Benjamins, Amsterdam, 65-80.
- Ozubko, J. D., & Fugelsang, J. (2011): «Remembering makes evidence compelling: Retrieval from memory can give rise to the illusion of truth», *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 37(1), 270-276.
- Parret, H. (1978): «Éléments d'une analyse philosophique de la manipulation et du mensonge. Prépublication», *Documents de Travail et Prépublications. Centro Internazionale di Semiotica e di Linguistica. Università di Urbino Urbino* 70 , 1-33.
- Pohl, R. (éd.), (2004): *Cognitive Illusions: a handbook on fallacies and biases in thinking, judgment and memory*. Psychology Press, New York.
- Reder, L. M. & Kusbit, G. W. (1991): «Locus of the Moses illusion: Imperfect encoding, retrieval, or match?», *Journal of Memory and Language* 29, 633-654.
- Rigotti, E. (2005): «Towards a typology of manipulative processes», in de Saussure, L. & Schulz, P (éds), *Manipulation and Ideologies in the Twentieth Century: Discourse, language, mind*. John Benjamins, Amsterdam, 61-83.
- Saussure, L. de (2005): «Manipulation and cognitive pragmatics: preliminary hypotheses», in de Saussure, L. & Schulz, P (éds), *Manipulation and Ideologies in the Twentieth Century: Discourse, language, mind*. John Benjamins, Amsterdam, 113-145.
- Saussure, L. de & Schulz, P. (éds) (2005): *Manipulation and Ideologies in the Twentieth Century: Discourse, language, mind*. John Benjamins, Amsterdam.
- Sperber, D., Cara, F, & Girotto, V. (1995): «Relevance Theory explains the Selection Task», *Cognition* 57, 31-95.

- Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G. & Wilson, D. (2010), «Epistemic Vigilance», *Mind & Language*, 25: 359–393.
- Sperber, D. & Wilson, D. (1995): *Relevance: Communication & Cognition* (2<sup>ème</sup> éd.), Blackwell, Oxford.
- Van Dijk, T. (2006): «Discourse and manipulation», *Discourse & Society* 17 (3), 359-383.
- Wilson, D. & Sperber, D. (2012): *Meaning and Relevance*. Cambridge University Press, Cambridge / New York.
- Zajonc, R. B. (1968): «Attitudinal effects of Mere Exposure», *Journal of Personality and Social Psychology*, 9(2), 1–27.

# Biased argumentation and critical thinking

Vasco CORREIA, Universidade Nova de Lisboa

*A man with a conviction is a hard man to change. Tell him you disagree and he turns away. Show him facts or figures and he questions your sources. Appeal to logic and he fails to see your point.*

Leon Festinger (2008 [1956]: 3)

## 1. Introduction

Although the problem of biased argumentation is sometimes reduced to the problem of intentional biases (sophistry, propaganda, deceptive persuasion)<sup>1</sup>, empirical data on human inference is consistent with the view that people are often biased not because they want to, but because their emotions and interests insidiously affect their reasoning (Kunda 1990, Baron 1988, Gilovich 1991). Walton (2011: 380) highlights this aspect:

Many fallacies are committed because the proponent has such strong interests at stake in putting forward a particular argument, or is so fanatically committed to the position advocated by the argument, that she is blind to weaknesses in it that would be apparent to others not so committed.

This phenomenon is known as «motivated reasoning» and typically occurs unintentionally, without the arguer's awareness (Mercier & Sperber 2011: 58, Pohl 2004: 2). For example, a lawyer may be biased in the defense of a client because (s)he deliberately intends to manipulate the jury, to be sure, but also because the desire to win the case (or the sympathy toward the client, etc.) unconsciously distorts the way (s)he reasons and processes the relevant evidence. In such cases, the arguer is sincerely convinced that his or her

1 See for example Herman & Chomsky (1988), Praktanis & Aronson (1991), Walton (2006).

arguments are fair and reasonable, while in fact they are tendentious and fallacious.

In the past decades, empirical research in psychology and neurosciences not only confirmed that emotions greatly affect human reasoning, but also revealed to what extent and in which variety of ways this seems to happen. Although the significance of these studies has been questioned by some authors,<sup>2</sup> the dominant view is that people tend to fall prey to a host of cognitive and motivational biases that affect their inferential and judgmental reasoning. As Larrick (2004: 316) observes, «the existence of systematic biases is now largely accepted by decision researchers, and, increasingly, by researchers in other disciplines». At any rate, psychologists now investigate dozens of types of cognitive illusions (Pohl 2004: 1, Thagard 2011: 164).

The phenomenon of motivated reasoning poses a considerable challenge for normative theories of argumentation, which tend to assume that the rules of logic and dialectic are sufficient to ensure the reasonableness of people's arguments. Insofar as motivational biases tend to occur unconsciously, it appears that even well-intended arguers, who genuinely wish to reason in fair terms, may end up putting forward arguments that are skewed and tendentious. To that extent, the intentional effort to observe the rules of argumentation may not suffice to ensure the rationality of debates. As Thagard (2011: 157) points out, «it would be pointless to try to capture these [motivated] inferences by obviously fallacious arguments, because people are rarely consciously aware of the biases that result from their motivations». Moreover, this difficulty is aggravated by the fact that arguers often tend to

2 Some researchers argue that the discrepancies between normative models of rationality and people's reasoning are not indicative of human irrationality, but rather the result of (1) random performance errors, (2) computational limitations of the human brain, (3) a misconception of the relevant normative standards of rationality, and (4) a different interpretation of the task by the subject (for a review, see Stanovich & West 2000). Gigerenzer (2008: 13), in particular, suggests that cognitive illusions are in fact adaptive forms of reasoning which promote the achievement of goals under constraints of time and information. Cohen (1981: 317), on the other hand, insists on the shortcomings of some empirical studies, which, in his view, neglect the differences of interpretation between the experimenters and the subjects, and involve mental tasks that are not representative of the normal conditions of reasoning. Other researchers have also questioned the presumed universality of such results, which are based almost exclusively on samples drawn from western, educated populations (Heinrich *et al.* 2010).

rationalize their biases; in other words, come up with good ‘reasons’ to justify *post-factum* beliefs initially acquired under the influence of motives (desires, goals, emotions).

This chapter has two purposes. The first is to elucidate some of the ways in which motivational biases lead arguers to commit unintentional fallacies. Drawing on recent work in psychology and argumentation theory, I explore the hypothesis that there are privileged links between specific motivational biases and specific forms of fallacious reasoning. To make this point clearer, I propose to categorize motivational biases in three different classes, according to the type of motive that underlies them: (1) *wishful thinking*, in which people are led to believe that p because they desire that p, (2) *aversive thinking*, in which people are led to believe that p because of the anxiety that not-p, and (3) *fretful thinking*, in which, as Thagard (2011: 159) explains, «people believe something, not just despite the fact that they fear it to be true, but partly because they fear it to be true».

The second purpose of this chapter is to argue that, even though motivational biases are in principle unintentional, there are certain control procedures that arguers can adopt if they wish to counteract their error tendencies. I identify several ‘debiasing strategies’ and argue that they can significantly contribute to promote the rationality of people’s argumentative reasoning, both at a dialogical and at an individual level. While most normative theories of argumentation focus on the problem of establishing the ideal rules of how people ought to argue, the last section of this article focuses on the problem of what discussants can do to effectively adjust their behavior to those rules.

The proposed categorization of motivational biases provides a structure for the three first sections of this chapter. In section 2, I examine the effects of wishful thinking in everyday discourse and suggest that it underlies our tendency both to commit the *argumentum ad consequentiam* fallacy and to fall prey to the ‘confirmation bias’, which in turn tends to accentuate the problem of polarization of opinions. In section 3, I argue that there is an intimate correlation between aversive thinking and two forms of fallacious reasoning: ‘misidentifying the cause’, on the one hand, and ‘slothful induction’, on the other. Furthermore, I explore Festinger’s (1957) hypothesis that aversive thinkers tend to become ‘defensive’ and to explain away their inconsistencies through rationalizing (rather than rational) arguments. In section 4, I examine

the puzzling phenomenon of ‘fretful thinking’ – or ‘fear-driven inference’ (Thagard 2011: 159) – in which arguers are biased in a self-defeating way, typically by ‘jumping to (negative) conclusions’ or by committing the ‘slippery slope’ fallacy. Finally, in section 5, I briefly examine several debiasing strategies and show how they can be useful to counteract the effects of biases upon people’s everyday reasoning.

## 2. Wishful thinking

Wishful thinking is generally described as a form of motivated reasoning in which the subject is led to conclude that  $p$  under the influence of a desire that  $p$ . Although this phenomenon is sometimes reduced to the inference ‘I wish that  $p$ , therefore  $p$ ’, it seems doubtful that wishful thinkers actually commit the fallacy in those terms. As much as I may wish to be beautiful, rich and famous, for example, it is clear that the inference ‘I wish to be beautiful, rich and famous, therefore I am beautiful, rich and famous’ is very unlikely to persuade me.<sup>3</sup>

In most cases of wishful thinking, arguers are not aware that they reach the conclusion that  $p$  merely because they desire that  $p$ . Instead, desire-driven inferences seem to involve a more complex and indirect type of fallacy – namely, a fallacious version of the argument from consequences: ‘If  $p$  than  $q$ . I wish that  $q$ . Therefore  $p$ ’. As Walton (2006: 106) observes, the argument from consequences may or may not be fallacious, depending essentially on three critical questions: (1) how likely it is that the consequence will follow, (2) what evidence is provided to support the claim that the consequence will follow, and (3) whether there are consequences of the opposite value that ought to be taken into account. If the arguer’s claim is based merely on the desirability of the consequence, regardless of its likelihood, his or her argument is probably fallacious. Consider the following example: «Of course

3 Most philosophers and psychologists agree that it is impossible to decide to believe something *hic et nunc* (direct doxastic voluntarism), although it may be possible to control indirectly certain beliefs, for example via a selective exposure to the available evidence (see Mele 2001b, for a review).

the environment talks will succeed. Otherwise it means mankind is on the way out» (Pirie 2006: 176). Although the mutual desire to avoid a catastrophic situation is a valid reason to believe that the parties involved in the talks will reach an agreement, this reasoning seems fallacious insofar as the undesirable consequences of a failure in the environment talks cannot by themselves guarantee that they will succeed (other factors, such as economic interests, may prevail).

Mele (2001a: 87) observes that desires tend to induce irrational reasoning indirectly, by affecting the subject's attention to the available evidence:

Data that count in favor of (the truth of) a hypothesis that one would like to be true may be rendered more vivid or salient given one's recognition that they so count; and vivid or salient data, given that they are more likely to be recalled, tend to be more 'available' than pallid counterparts.

This helps explain why wishful thinking is often associated with the well-documented confirmation bias, which consists precisely in the tendency to search for evidence that supports what we already believe in, or what we want to be true (Baron 1988: 280, Oswald & Grosjean 2004: 79). Thus, for example, the desire that my philosophical position is correct may surreptitiously lead me to focus too much on sources that seemingly confirm it, and not enough on sources that seemingly disconfirm it. Likewise, «people who want to believe that they will be academically successful may recall more of their past successes than of their failures» (Kunda 1990: 483). In a classic experiment, Lord *et al.* (1979) were able to demonstrate that this bias tends to aggravate the phenomenon of 'attitude polarization' even when people are exposed to the same body of information. The researchers exposed subjects supporting and opposing the death penalty to descriptions of two fake studies, one confirming and one disconfirming the hypothesis that capital punishment deters violent crime. Predictably, they found that both proponents and opponents of the death penalty rated the study that confirmed their own views as more convincing and probative. Less predictably, though, they found that the 'pro' subjects became even more favorable to the capital punishment after being exposed to the information, and that the 'anti' subjects became even more opposed to it. In other words, the polarization of opinions seemed to have increased after exposure to information, despite the fact that the information was the same.

Assuming that this phenomenon is caused by unconscious biases, it becomes all the more difficult to prevent it. People who succumb to the confirmation bias do not try to find support for their preexisting beliefs by deliberately twisting or misinterpreting the available evidence. As Kunda (1990: 494) explains, the problem is rather that «cognitive processes are structured in such a way that they inevitably lead to confirmation of hypotheses». The classic assumption to explain such tendencies is that people are generally motivated to protect their belief system from potential challenges (Albarracín & Vargas 2009, Festinger *et al.* 2008 [1956], Mercier & Sperber 2011). Festinger *et al.* (2008 [1956]: 3) write: «We are familiar with the variety of ingenious defenses with which people protect their convictions, managing to keep them unscathed through the most devastating attacks». Interestingly enough, some studies indicate that people who are confident about the resilience of their beliefs are more willing to examine evidence that contradicts them, and, conversely, that people who are doubtful about their ability to defend their beliefs from future challenges tend to prefer exposure to information consistent with them (Albarracín & Mitchell 2004). According to Mercier and Sperber (2011: 65) the confirmation bias helps arguers meet the challenges of others and even contributes to a prosperous division of cognitive labor, «given that each participant in a discussion is often in a better position to look for arguments in favor of his or her favored solution (situations of asymmetric information)».

The problem, however, is that this tendency to gravitate toward information that justifies our preexisting opinions leads to a polarization of opinions that is arguably detrimental to the purpose of debates. Given the same body of evidence, people with different opinions will tend to focus on elements that are susceptible to cause their views to move even further apart. Lord *et al.* (1979: 2108) vehemently stress that point: «If our study demonstrates anything, it surely demonstrates that social scientists cannot expect rationality, enlightenment, and consensus about policy to emerge from their attempts to furnish 'objective' data about burning social issues». In addition, this aspect seems to be enhanced by the way people tend to consume information nowadays, as Mooney (2011: 3) observes, «through the Facebook list of friends, or tweets that lack nuance or context, or narrowcast and often highly ideological media that have small, like-minded audiences».



### 3. Aversive thinking

Whereas in wishful thinking arguers unduly infer that *p* is true because of the desire that *p*, in aversive thinking arguers unduly infer that *p* because of the anxiety that not-*p*. Although the anxiety that not-*p* is generally accompanied by a correlative desire that *p*, the two affects seem to be causally independent and may trigger different forms of motivated reasoning (Barnes 1997: 52, Johnston 1989: 72). Like wishful thinking, aversive thinking is considered to be a ‘positive’ illusion inasmuch as it yields a significant psychological gain, namely: the reduction of the subject’s anxiety.

Aversive thinking tends to arise when arguers are confronted with evidence suggesting that what they fear might be true. For example, a man who is diagnosed with a terminal illness may reject the doctors’ arguments and persist in believing that he will survive. His way of reasoning is presumably constrained by the anxiety of thinking that his days are numbered, which can of course be psychologically devastating. But aversive thinking need not be so extreme. In everyday debates, people often become ‘defensive’ simply because one of their convictions is being challenged. As Johnson & Blair (1983: 193) observe, this seems to happen in virtue of the arguer’s ‘egocentric commitment’ to a given standpoint, as when people are blinded by their attachment to an ideology, a group or an institution. According to Sherman & Cohen (2002: 120) such defensive responses stem more fundamentally from a motivation to protect self-worth and the integrity of the self: «Because the motivation to maintain self-worth can be so powerful, people may resist information that could ultimately improve the quality of their decisions».

One of the most effective forms of aversive thinking is rationalization, i.e., the effort to justify an irrational attitude by invoking ‘good’ reasons instead of the true reason. It is notoriously difficult to refute the arguments of a person who rationalizes, given that the reasons (s)he invokes are not necessarily false. Someone addicted to pills, for example, may be able to provide seemingly reasonable explanations for abusing medication (stress at work, domestic problems, headaches, sleeping disturbances, etc.). Yet, much like the alcoholic who claims to drink for a reason, (s)he will refuse to acknowledge that (s)he takes pills mainly because of a drug addiction. Given

the social stigma associated to that diagnosis, the motivation to deny it can be strong enough to distort the way (s)he reasons.

In many cases, rationalization leads arguers to commit the fallacy of misidentifying the cause, which according to Tindale (2007: 179) may take two forms: «In the first instance, we may falsely identify X as the cause of Y when on closer inspection a third factor, Z, is the cause of both X and Y. In the second case, we may confuse a cause and an effect: identifying X as the cause of Y when it is actually Y that causes X». The latter case is illustrated precisely by the addicted person who claims that (s)he takes drugs because of all sorts of problems (work, family, health, etc.), when in general those problems are already a consequence of the abuse of drugs (Twerski 1997: 34).

According to the theory of cognitive dissonance (Festinger 1957, Aronson 1969) such responses arise when the person holds two or more ‘cognitions’ (ideas, beliefs, opinions) that are psychologically inconsistent with each other. Inasmuch as the occurrence of dissonance admittedly produces anxiety and psychological discomfort, individuals strive toward consistency within themselves by rationalizing one of the cognitions in question. Hence, Festinger (1957: 3) writes:

The person who continues to smoke, knowing that it is bad for his health, may also feel (a) he enjoys smoking so much it is worth it; (b) the chances of his health suffering are not as serious as some would make out; (c) he can’t always avoid every dangerous contingency and still live; and (d) perhaps even if he stopped smoking he would put on weight which is equally bad for his health. So, continuing to smoke is, after all, consistent with his ideas about smoking.

When the attempt to explain away the inconsistency is successful, dissonance is reduced and so is the anxiety associated to it<sup>4</sup>. Several experiments confirm that people tend to rationalize their inconsistencies (for a review, see Albarracín & Vargas 2009). In a classic study, Festinger & Carlsmith’s (1959)

4 For one reason or another, however, attempts to achieve consistency may fail and the psychological discomfort persists. Moreover, even successful rationalizations can lead to more anxiety in the long term, as Barnes (1997: 35) points out: “The reduction of anxiety can lead to other anxieties, sometimes far greater ones. The gain, therefore, is not necessarily an all-things-considered gain, nor is it necessarily beneficial for the person”.

asked students to work for an hour on boring tasks such as turning pegs a quarter turn over and over again. Participants were then asked to convince another student that the tedious and monotonous tasks were actually enjoyable and exciting. While some of the participants were paid \$20 for doing this, others were paid merely \$1. Surprisingly, when the participants were asked how much they really enjoyed performing the tasks, those who were paid \$1 rated the tasks as more enjoyable than those who were paid \$20. The researchers speculated that all the participants experienced dissonance between the conflicting cognitions: 'The tasks were tedious' and 'I told someone that the tasks were exciting'. However, those who were paid \$20 had a great deal of justification for lying to the other student, and therefore experienced less dissonance. Those who were paid \$1, on the other hand, experienced a greater need to justify their action and presumably persuaded themselves that they really believed what they said.

This tendency to rationalize seems to be particularly strong when the perceived inconsistencies are liable to threaten the arguer's emotional attachment to the standpoint. In a recent study, Westen *et al.* (2006) used functional neuroimaging to test motivated reasoning on political partisans during the U.S. Presidential election of 2004. The subjects were shown a set of slides presenting contradictory pairs of statements either from their preferred candidate, from the opposing candidate, or from a neutral figure. In addition, one of the slides presented an exculpatory statement that explained away the apparent contradiction. Then they were asked to consider whether each candidate's statements were inconsistent or not. Predictably, the subject's ratings provided strong evidence of motivated reasoning. First, they were substantially more likely to evaluate as inconsistent statements made by the candidate they opposed. And second, they were much more likely to accept the exculpatory statements for their own candidate than those for the opposing candidate. In addition, the scanners revealed that the brain regions specifically involved in emotion processing were strongly activated when the subjects evaluated contradictory statements by their preferred candidate, but not when they evaluated the other figure's contradictions. The researchers concluded that biases were due to the participants' effort to reduce cognitive dissonance: «Consistent with prior studies of partisan biases and motivated reasoning, when confronted with information about their

candidate that would logically lead to an emotionally aversive conclusion, partisans arrived at an alternative conclusion» (Westen *et al.* 2006: 1955).

These results suggest that, in many cases of aversive thinking, people's arguments are less the reason why they adhere to a given belief than the *post factum* justification of their preexisting belief. Thus, for example, a person may believe in the immortality of the soul in virtue of a religious education, and nonetheless invoke apparently reasonable arguments to support that conviction, as if those arguments were the reason why (s)he held that belief in the first place. To paraphrase one of Aldous Huxley's famous quotes in *Brave New World*, it seems fair to say that people often come up with seemingly good reasons to justify beliefs that they initially acquired for bad (or unjustified) reasons. This constitutes, according to Haidt (2010: 355), the fundamental «Problem of Motivated Reasoning: The reasoning process is more like a lawyer defending a client than a judge or scientist seeking the truth».

In more extreme cases of aversive thinking, such as denial, the anxiety toward the undesired conclusion is so intolerable that the subject rejects it in the teeth of evidence. As Mele (1982) suggests, this is possible because psychological inferences are not as compulsory as logical deductions: «One may believe that p is true and that p entails q without believing that q is true; and this is the kind of thing that may be explained by a want, fear, or aversion of the person». In such cases, it all happens as though the subject accepted the premises of the reasoning but not the conclusion that follows, presumably because the anxiety somehow 'inhibits' the inferential step. The terminally ill patient who refuses to accept his diagnosis, for example, may acknowledge that the exams are reliable and that the doctors are competent but refuse nonetheless to adhere to what they clearly suggest. Sartre (1943: 100) puts forward a similar hypothesis in the famous interpretation of the homosexual who denies his sexual orientation: «[he] acknowledges all the elements that are imputed to him and yet refuses to draw the obvious conclusion».

Such cases of aversive thinking seem to involve a specific form of *ignoratio elenchi* that Schopenhauer (1831: 11) called «denial of conclusion, *per negationem consequentiae*», which some informal logicians now term slothful induction, i.e., «the mistake of underrating the degree of probability with which a conclusion follows from evidence» (Baker 2003: 264). In a sense, the fallacy of slothful

induction appears to be the symmetric opposite of the fallacy of hasty generalization, insofar as the later involves ‘jumping to conclusions’ on the basis of insufficient evidence, whereas the former involves failing to draw a conclusion in the face of sufficient evidence (Correia 2011: 120). In some cases this may occur indirectly: The subject appreciates the evidence that  $p$  but rejects the very link between  $p$  and the undesired conclusion  $q$ . When the motivation to deny that  $q$  is strong enough, Thagard (2011: 155) writes, «you need to question your belief in if  $p$  then  $q$  and  $p$ , rather than blithely inferring  $q$ ». To return to the case of the terminally ill patient, it may happen for example that he challenges the predictions of standard medicine and turns to less warranted therapeutic methods in a desperate attempt to deny the imminence of his death.

That being said, it is clear that people do not necessarily engage in aversive thinking whenever they feel reluctant to accept a certain reality. Whether the subject falls prey to a motivated illusion or not seems to depend on at least two factors: on the one hand, the degree of emotional attachment to the belief in question, and, on the other hand, the degree of reliability of the subject’s habits of thinking. I will return to this question in the last section.

#### 4. Fretful thinking

The phenomenon known as ‘fretful thinking’ (Beyer 1998: 108), ‘counterwishful thinking’ (Elster 2007: 384) and ‘twisted self-deception’ (Mele 2001b: 94) is surely the most puzzling form of motivated reasoning. Unlike wishful and aversive thinking, which both tend to create biases that are consistent with the individual’s goals, fretful thinking seems paradoxical in that it engenders self-defeating biases that yield unwelcome beliefs. This is what happens, for example, when a jealous husband is biased into thinking that his wife is having an affair, despite his not wanting it to be the case. Likewise, a pessimistic woman may underestimate her chances of getting a job despite her desire to get the job. In this type of case, it all happens as though the person’s fear that  $p$  somehow caused her to acquire the unwarranted belief that  $p$ . To that extent, as Thagard (2011: 159) points out,

«fear-driven inference is doubly irrational, from both a practical and theoretical perspective, because it gives the thinker unhappiness as well as erroneous beliefs». The author considers several plausible examples of this:

- (a) My lover looks distant, so he/she must be having an affair.
- (b) I haven't heard from my teenager for a few hours, so he's probably in trouble.
- (c) This rash means I have leprosy or some other serious disease.
- (d) The editor's delay in responding to my article means he/she hates it.

As it appears from these examples, fretful thinking often leads arguers to jump to (negative) conclusions without sufficient evidence. In general, this seems to happen because the subject focuses too much on the negative aspects of the issue, presumably due to the influence of a negative emotion (fear, jealousy, anxiety, etc.) on the way she processes information. Regarding the case of the jealous man, for example, it seems reasonable to suggest that «[he] believes that his wife is unfaithful because of the effects of his jealousy on the salience of his evidence or on the focus of his attention» (Mele 2001b: 101). Rather than considering indiscriminately all the relevant evidence, he tends to focus exclusively on the elements that seem to confirm his worst suspicions, and subsequently falls into the illusion that his wife is (probably) cheating on him. This aspect is perhaps more obvious in pathological cases of 'morbid jealousy' (or 'Othello Syndrome'), in which individuals incessantly accuse their partner of infidelity «based on incorrect inferences supported by small bits of 'evidence' (e.g., disarrayed clothing or spots on the sheets), which are collected and used to justify the delusion» (American Psychiatric Association 2000: 325). A similar explanation plausibly accounts for the case of the parents who jump to conclusions regarding their teenager's safety or the hypochondriac who panics because of a mere rash: a negative emotion leads them to contemplate uniquely the negative aspect of things and the arguments they put forth tend to be affected by a pessimism bias.

In such cases arguers seem to fall prey to what Schkade & Kahneman (1998: 340) call the focusing illusion: «When a judgment about an entire object or category is made with attention focused on a subset of that category, a focusing illusion is likely to occur, whereby the attended subset is overweighed relative to the unattended subset». While it may not be

fallacious, strictly speaking, to confine one's reasoning solely and exclusively to the negative side of things, such a tendentious interpretation of the available evidence is likely to undermine both the rationality and the credibility of the resulting arguments. Walton also highlights this point:

An argument is more plausible if it is based on a consideration of all the evidence in a case, on both sides of the issue, than if it is pushing only for one side and ignoring all the evidence, even if it may be good evidence, on the other side. So if an argument is biased, that is, if it pushes only for one side, we discount that argument as being worthless. (Walton: 2006: 238)

There also seems to be a privileged link between the phenomenon of fretful thinking and the fallacy of slippery slope, given that the latter typically leads the arguer to draw a dreadful conclusion from a somewhat dubious causal association between events. As Walton (2006: 107) observes, the slippery slope leads the arguer to predict a «particularly horrible outcome [which] is the final event in the sequence and represents something that would very definitely go against goals that are important for the participant...» In fact, most slippery slopes seem to be fear-driven inferences that alert to catastrophic and exaggerated scenarios on the basis of insufficient evidence: e.g., that usage of cannabis is the first step to the use of harder drugs; that immigration leads to the loss of traditional values and eventually to the loss of a national identity; that China's economic growth will lead to a military supremacy, which in turn will cause the decline of western powers; and so forth. It is difficult not to speculate that, in such cases, the propensity to commit the slippery slope is motivated by the arguer's fears (or 'fear-driven', as Thagard says), exactly as in the case of the jealous husband and in the case of the hypochondriac. Another plausible example would be the slippery slope motivated by xenophobic fears, as in the following example (Pirie 2006: 152): «If we allow French ideas on food to influence us, we'll soon be eating nothing but snails and garlic and teaching our children to sing the Marseillaise». Some authors hypothesize that, on such occasions, the person's reasoning is biased by an 'irrational emotion', i.e., an emotion that is either based on an irrational belief or not based on any belief at all (De Sousa 1987: 197, Elster 1999: 312). This hypothesis is consistent with the claim that negative illusions may have been beneficial in the evolutionary past, given that the tendency to assume the worst seems to encourage risk avoidance (Andrews & Thomson 2009). It seems plausible, for example, that delusional

jealousy might have increased people's vigilance against potential rivals, thereby discouraging infidelity between partners. Yet, in modern environments such biases lead to unfair and counter-productive arguments which seem to compromise the individual's goals, particularly when they are motivated by irrational attitudes – not just jealousy, but excessive jealousy, not just distrust, but unjustified distrust; not just pessimism, but unrealistic pessimism.

## 5. Critical thinking and argumentative self-regulation

There has been much controversy over whether motivational biases are inherent defects of human reason that tend to undermine the rationality of people's reasoning (Kahneman 2011, Kunda 1990, Gilovich 1991) or, on the contrary, adaptive mechanisms that tend to maximize decision-making under constraints of time and knowledge (Gigerenzer 2008, McKay and Dennett 2009, Taylor & Brown 1988). On the one hand, it seems plausible that motivational biases may turn out to be beneficial in light of the subject's environment and goals. From this perspective, Gigerenzer (2008:13) argues, «what appears to be a fallacy can often also be seen as adaptive behavior». Thus, for example, some studies indicate that wishful thinking and self-serving biases tend to enhance people's motivation, productivity and mood (Taylor & Brown 1988). On the other hand, however, we have seen that biases may also lead to maladaptive responses, such as denial, risk-mismanagement, prejudice, polarization of opinions, and rationalization (for a review, see Dunning *et al.* 2004). Furthermore, even assuming that biases may be adaptive from a *utilitarian* standpoint, it is clear that they often compromise the rationality of people's arguments from a *dialectical* standpoint. As Johnson and Blair (2006: 191) observe, motivational biases stem from egocentric and emotional attachments which «often result in a failure to recognize another point of view, to see the possibility of an objection to one's point of view, or to look at an issue from someone else's point of view».



Be that as it may, for the purpose of this section it is enough to assume that motivational biases may *sometimes* lead to irrational attitudes in everyday contexts of argumentation. The question to be asked, then, is whether arguers should do something to minimize their irrational tendencies. A number of virtue epistemologists and belief theorists have recently argued that, even though motivational biases are typically unintentional, subjects have the ‘epistemic obligation’ to try to mitigate the effects of biases upon their cognitive processes (Adler 2002, Audi 2008, Engel 2000, Mele 2001a). I have tried to defend elsewhere that this claim is also pertinent in the realm of argumentation theory (Correia 2012), and, more specifically, that the effort to counteract one’s motivational biases should be included in what Johnson (2000: 165) calls the arguers’ «dialectical obligations». For the present purpose, however, I will confine my analysis to the descriptive question that seems to be presupposed by the normative one: Assuming that arguers have the dialectical obligation to debias themselves, how can they achieve this? After all, it only makes sense to suggest that discussants are partly responsible for their irrational thinking if there is something they can do to prevent it.

Perelman and Olbrechts-Tyteca (1969: 119) maintain that biases are unavoidable flaws that are inherent in the process of argumentation: «All argumentation is selective. It chooses the elements and the method of making them present. By doing so it cannot avoid being open to accusations of incompleteness and hence of partiality and tendentiousness». To some extent, at least, the authors are probably right, for it is virtually impossible not to let emotions influence the way we reason in one way or another; and, perhaps for that reason, it is almost a commonplace to acknowledge that the ideal of impartiality is unachievable. Having said this, it is important to bear in mind that «rationality is a matter of degree», as Baron (1988: 36) points out, and that arguers may at least try to minimize the phenomenon of motivated reasoning.

As a matter of fact, it appears that arguers are not condemned to remain the helpless victims of their error tendencies. Even though motivational biases are typically unconscious, there are certain control strategies that arguers can adopt if they wish to counteract the effects of biases upon their reasoning, both at an individual and at a dialogical level. In particular, arguers may adopt a certain number of ‘debiasing strategies’ designed to promote the rationality of their attitudes in a debate.

Before examining some of these strategies, it is worth noting that the very awareness of our biases can perhaps contribute to mitigate their effects. Those who are «open-minded enough to acknowledge the limits of open-mindedness», as Tetlock (2005: 189) elegantly puts it, seem to be in a better position to overcome their cognitive weaknesses and to ensure the rationality of their arguments. For example, a scientist who is aware of the heuristic distortions induced by the confirmation bias may attempt to offset them by forcing herself to examine thoroughly sources that seem to contradict her position. Likewise, arguers who accept the notion that they may be biased without being aware of it are perhaps more likely to remain vigilant against such biases, and perhaps more willing to meet their opponents halfway in the process of solving a difference of opinion. Hence, Thagard (2011: 160) writes, «critical thinking can be improved, one hopes, by increasing awareness of the emotional roots of many inferences».

Second, arguers who wish to make sure that their arguments are fair and balanced may adopt the strategy of «playing the devil’s advocate» (Stuart Mill 1859: 35), i.e., «throw themselves into the mental position of those who think differently from them». According to Johnson (2000: 170) the effort to examine the set of standard objections to our own views constitutes a dialectical obligation that arguers must fulfill even in the absence of an actual opponent<sup>5</sup>. This normative requirement seems particularly useful to counteract the confirmation bias and the fallacy of cherry picking, since it exhorts people to contemplate alternative standpoints and sources of information which they spontaneously might tend to neglect. As Larrick (2004: 323) explains, «the strategy is effective because it directly counteracts the basic problem of association-based processes – an overly narrow sample of evidence – by expanding the sample and making it more representative».

Third, biases can more easily be detected if discussants proceed to the ‘analytic reconstruction’ of arguments (Walton 1989b: 170, Eemeren & Grootendorst 2004: 95). By analyzing their discourse into its elementary components, arguers have a better chance to detect hidden biases and to externalize their implicit commitments. Walton (2006: 227-228) stresses that

5 Johnson (2000: 165) observes that traditional approaches have focused too much on what he calls the ‘illative core’ of arguments, i.e., the set of premises that arguers advance in support of the conclusion, and not enough on the ‘dialectical tier’, i.e., the set of alternative positions and plausible objections that must be addressed.

a «bias may not only be hidden in the emotive words used to make a claim, it may also be hidden because the claim itself is not even stated, only implied by what was not said». Oftentimes, arguers themselves are unaware of their ‘dark-side commitments’ and of the extent to which these can bear on their reasoning. For example, a person’s belief in the existence of God may lead her to bring forward arguments that inadvertently beg the question with regard to matters such as morality and politics. The effort to analyze the components of argumentative discourse seems to contribute to render such commitments explicit, thereby allowing discussants to become aware of their biases.

More generally, arguers may promote critical thinking by improving their argumentative skills. After all, people who have a good understanding of the rules of logic, statistics and argumentation are presumably more likely to detect their own fallacies. Tversky and Kahneman (2008) were able to confirm this hypothesis in a recent replication of the well-known ‘Linda problem’. In the original versions of the experiment (Tversky & Kahneman 1983) the researchers submitted to undergraduates a description of Linda, a fictitious person, as a thirty-one years old activist deeply concerned with issues of discrimination and social justice. Then they asked the participants which of the following possibilities is more likely: (A) Linda is a bank teller, or (B) Linda is a bank teller and is active in the feminist movement? Surprisingly, about 85% to 90% of undergraduates at several major universities chose the second option, thereby transgressing an elementary rule of probabilities: The conjunction of two events cannot be more probable than one of the events alone. Yet, a more recent version of the experiment (Tversky & Kahneman 2008: 120), conducted with graduate students with statistic education, revealed that only 36% committed the fallacy, which seems to indicate that, at least in certain cases, the development of deductive skills can work as a safeguard against systematic errors of intuitive reasoning<sup>6</sup>.

That is not to say that deductive skills alone suffice to ensure the rationality of the arguer’s attitudes in a debate. As Paul (1986: 379) rightly observes, «it is possible to develop extensive skills in argument analysis and construction without ever seriously applying those skills in a self-critical way

6 Cohen (1981) would object that subjects with training in logic, probability theory and statistics only appear to be better intuitive reasoners because these are precisely the tasks that they are trained to perform.

to one's own deepest beliefs, values, and convictions». Some authors suggest that the reasonableness of debates depends just as much, if not more, upon the discussants' argumentational virtues, that is, on the set of dispositions and character traits that tend to promote good thinking (Aberdein 2010: 169, Cohen 2009: 49). For example, virtues such as open-mindedness, fairness, intellectual honesty, perseverance, diligence and humility seem to offset many of the biasing tendencies examined earlier. The advantage of fostering such virtues is that they tend to form a sort of 'second nature' (Montaigne 1967 [1580]: 407, Ryle 1949: 42) which enables people to reason in fair terms almost spontaneously, without a permanent effort to remain impartial.

Finally, discussants have the possibility of adopting what decision-theorists call 'precommitment strategies' of self-control, which may be described as self-imposed constraints designed to avoid irrational attitudes (Elster 2007, Loewenstein *et al.* 2003). In what regards argumentation contexts, such constraints aim at regulating the conditions under which the information is processed and the arguments set out. Thus, a scientist who is about to submit an article on the issue of global warming, but recognizes that her convictions are susceptible to bias her analysis, may commit in advance to several control strategies: for example, verify that she did not overlook any disconfirming evidence (confirmation bias); ask a colleague to try to detect unintentional biases; carefully examine and respond to the standard set of counterarguments; make sure that these have not been misrepresented (straw man argument); and so forth. To be sure, it may not always be easy to adopt self-regulation strategies in everyday debates, given the usual constraints of time and information, but, as Kahneman (2011: 131) points out, «the chance to avoid a costly mistake is sometimes worth the effort».

## 6. Conclusion

This paper sought to elucidate the problem of how goals and emotions can influence people's reasoning in everyday debates. By distinguishing between three categories of motivational biases, we were able to see that arguers tend to engage in different forms of fallacious reasoning depending on the type of

motive that underlies their tendentiousness. We have examined some plausible connections between specific types of biases and specific types of fallacies, but many other correlations could in principle be found. Although psychology studies consistently confirm people's propensity to be biased, motivated fallacies often appear persuasive and difficult to detect because of the arguers' tendency to rationalize their inconsistencies (Festinger 1957) and because of the 'illusion of objectivity' (Kunda 1990: 483) that results from it. Given that these processes tend to occur unconsciously, people's intentional efforts to observe the rules of argumentation are not always sufficient to prevent them from being biased.

Yet argumentational biases are not inevitable and arguers can (and perhaps ought to) counteract their irrational attitudes by submitting the process of argument-making to indirect strategies of control. The aim of argumentative self-regulation is to make sure that arguers effectively observe the rules of critical discussion in real-life contexts. In my view, this effort must be rooted in a good understanding of the very mechanisms that underlie our error tendencies. As Thagard (2011: 158, 164) suggests, «critical thinking requires a psychological understanding of motivated inference» and «a motivation to use what is known about cognitive and emotional processes to improve inferences about what to believe and what to do». The above-described strategies are mere examples of what arguers can do to promote the rationality of the way they reason, but there may be, in principle, as many debiasing strategies as there are types of motivated reasoning.

## References

- Aberdein, A. (2010): "Virtue in argument", *Argumentation* 24 (2), 165-179.
- Adler, J. (2002): *Beliefs Own Ethics*. Bradford, MIT, Cambridge MA.
- Albarracín, D. & Vargas, P. (2009): "Attitudes and persuasion: From biology to social responses to persuasive intent", in Fiske, S., Gilbert, D. & Lindzey, G. (eds.), *Handbook of Social Psychology*, Wiley & Sons, Hoboken NJ, 394-427.
- American Psychiatric Association (2000): *Diagnostic and Statistical Manual of Mental Disorders*. Fourth Edition, Text Revision, American Psychiatric Association, Washington, DC.
- Andrews, P. & Thomson, J. (2009): "The bright side of being blue", *Psychological Review* 116 (3), 620-654.

- Audi, R. (2008): "The ethics of belief: Doxastic self-control and intellectual virtue", *Synthese* 161, 403-418.
- Baker, S. (2003): *The Elements of Logic*. McGraw-Hill, New York.
- Barnes, A. (1997): *Seeing Through Self-Deception*. Cambridge University Press, Cambridge.
- Baron, J. (1988): *Thinking and Deciding*. Cambridge University Press, Cambridge.
- Beyer, L. (1998): "Keeping self-deception in perspective", in Dupuy, J.-P. (ed.), *Self-Deception and Paradoxes of Rationality*. CSLI Publications, 87-111.
- Cohen, D. (2009): "Keeping an open mind and having a sense of proportion as virtues in argumentation", *Cogency* 1 (2), 49-64.
- Cohen, J. (1981): "Can human irrationality be experimentally demonstrated?", *Behavioral and Brain Sciences* 4, 317-370.
- Correia, V. (2011): "Biases and fallacies: The role of motivated irrationality in fallacious reasoning", *Cogency* 3 (1), 107-126.
- , (2012) "The ethics of argumentation", *Informal Logic* 32 (2), 219-238.
- De Sousa, R. (1987): *The Rationality of Emotion*, M.I.T. Press.
- Dunning, D., Heath, C. & Suls, J. M. (2004): "Flawed self-assessment: Implications for health, education, and the workplace", *Psychological Science in the Public Interest* 5, 69-106.
- Elster, J. (2007): *Explaining Social Behavior*. Cambridge University Press, Cambridge.
- Engel, P. (2000): *Believing and Accepting*. Kluwer, Dordrecht.
- Festinger, L., Riecken, H. & Schachter, S. (2008) [1956]: *When Prophecy Fails*. Printer & Martin, London.
- Festinger, L. (1957): *A Theory of Cognitive Dissonance*. Stanford University Press, Stanford.
- Festinger, L. & Carlsmith, J.M. (1959): "Cognitive consequences of forced compliance", *Journal of Abnormal and Social Psychology* 58, 203-210.
- Gilovich, T. (1991): *How We Know What Isn't So*. The Free Press, New York.
- Gigerenzer, G. (2008): *Rationality for mortals*. Oxford University Press, New York.
- Haidt, J. (2010): "The emotional dog and its rational tail: A social intuitionist approach to moral judgment", in Nadelhoffer, T., Nahmias, E. & Nichols, S. (eds.), *Moral Psychology*. Wiley-Blackwell, West Sussex, 343-357.
- Heinrich, J., Heine, S. & Norenzayan, A. (2010): "The weirdest people in the world?", *Behavioral and Brain Sciences* 33, 61-135.
- Herman, E.S. & Chomsky, N. (1988): *Manufacturing Consent*. Pantheon Books, New York.
- Johnson, R. & Blair, J. (1983): *Logical Self-defense*. McGraw-Hill, Toronto.
- , (2000): *Manifest Rationality*. Lawrence Erlbaum, Mahwah, NJ.
- Johnston, M. (1989): "Self-deception and the nature of mind", in Rorty, A. & McLaughlin, B. (eds.), *Perspectives on Self-Deception*. University of California Press, Berkeley, 63-91.
- Kahneman, D., (2011): *Thinking, Fast and Slow*. Penguin Group, London.
- Kunda, Z. (1990): "The Case for Motivated Reasoning", *Psychological Bulletin* 108 (3), 480-498.
- Larrick, R. (2004): "Debiasing", in Koehler, D. & Harley, N. (eds.), *Blackwell Handbook of Judgment and Decision Making*. Blackwell Publishing, Wiley, 316-337.
- Loewenstein, G., Read, D. & Baumeister, R. (eds) (2003): *Time and Decision*. Russell Sage Foundation, New York.
- McKay, R. T. & Dennett, D. (2009): "The Evolution of Misbelief?", *Behavioral and Brain Sciences* 32, 493-561.

- Mele, A. (1982): "Self-deception, action and will: Comments", *Erkenntnis* 18, 159-164.
- , (2001a): *Autonomous Agents*. Oxford University Press, Oxford/New York
- , (2001b): *Self-Deception Unmasked*. Princeton University Press, Princeton.
- Mercier, H. & Sperber, D. (2011): "Why do humans reason? Arguments for an argumentative theory", *Behavioral and Brain Sciences* 34, 57-74.
- Montaigne, M. (1967) [1580]: *Essais*. Seuil, Paris.
- Mooney, C. (2011): *The science of why we don't believe in science*. Available at <http://www.motherjones.com/politics/2011/03/denial-science-chris-mooney>. Last accessed 21.01.2014.
- Oswald, M. & Grosjean, S. (2004): "Confirmation bias", in Pohl, R. (ed.), *Cognitive Illusions*. Psychology Press, Hove/New York, 79-96.
- Paul, W.R. (1986): "Critical thinking in the strong and the role of argumentation in everyday life", in van Eemeren, F., Grootendorst, R., Blair, A. & Willard, C. A. (eds.), *Argumentation*. Foris Publications, Dordrecht.
- Pirie, M. (2006): *How to Win Every Argument*. Continuum International Publishing Group, New York.
- Pohl, R. (ed.) (2004): *Cognitive Illusions*. Psychology Press, Hove/New York.
- Praktanis, A. & Aronson, E. (1991): *Age of Propaganda*. W. H. Freeman & Co.: New York.
- Ryle, G. (1949): *The Concept of Mind*. Penguin Books, New York.
- Sartre, J.-P. (1943): *L'être et le néant*. Seuil, Paris.
- Schkade, D. & Kahneman, D. (1998): "Does living in California make people happy?", *American Psychological Society* 9 (5), 340-346.
- Schopenhauer, A. (1831): *The Essays of Arthur Schopenhauer; The Art of Controversy*, transl. B Saunders, The Echo Library, Middlesex.
- Sherman, D.K. & Cohen, G.L. (2002): "Accepting threatening information: Self-affirmation and the reduction of defensive biases", *American Psychological Society* 11 (4), 119-123.
- Stanovich, K. & West, R. (2000): "Individual differences in reasoning: Implications for the rationality debate", *Behavioral and Brain Sciences*, 23, 645-726.
- Stuart Mill, J. (1859): *On Liberty*. Forgotten Books, Charleston.
- Taylor, S. E. & Brown, J. (1988): "Illusion and Well-Being: A Social Psychology Perspective on Mental Health", *Psychological Bulletin* 103 (2), 193-210.
- Tetlock, P. (2005): *Political Judgment*. Princeton University Press, Princeton.
- Thagard, P. (2011): "Critical thinking and informal logic: Neuropsychologic perspectives", *Informal Logic* 31 (3), 152-170.
- Tindale, C. (2007): *Fallacies and Argument Appraisal*. Cambridge University Press, Cambridge.
- Tversky, A. & Kahneman, D. (1983): "Extensional versus intuitive reasoning: the Conjunction Fallacy in probability judgment", *Psychological Review* 90 (4), 293-315.
- , (2008): "Extensional versus intuitive reasoning: the Conjunction Fallacy in probability judgment", in Adler, J. & Rips, L. (eds.), *Reasoning: Studies of Human Inference and its Foundations*. Cambridge University Press, Cambridge, 114-135.
- Twerski, A. (1997): *Addictive Thinking*. Hazelden, Center City, Minnesota.
- Walton, D. (1989b): "Dialogue theory for critical thinking", *Argumentation* 3, 169-184.
- , (2006): *Fundamentals of Critical Argumentation*. Cambridge University Press, Cambridge.
- , (2011): "Defeasible reasoning and informal fallacies", *Synthese* 179, 377-407.

Westen, D., Blagov, P, Harenski, K., Kilts, C. & Hamann, S (2006): “Neural basis of motivated reasoning”, *Journal of Cognitive Neuroscience* 18 (11), 1947-1958.



# Vers une naturalisation de la rhétorique?

## Problèmes épistémologiques

Emmanuelle DANBLON, Université Libre de Bruxelles

### 1. Introduction

Ce chapitre a l'ambition de défendre un projet de naturalisation de l'activité rhétorique en cherchant à surmonter deux écueils que cette démarche doit veiller à éviter:

- Le réductionnisme (naturaliser, c'est réduire au biologique)
- Le post-modernisme (naturaliser, c'est sortir de sa discipline et utiliser des concepts que l'on ne maîtrise pas ou, pire, que l'on utilise de façon métaphorique)

Pour illustrer ces deux dangers, j'aborderai de façon critique les points de vue respectifs de Dan Sperber et de Jacques Lacan. Il s'agit d'une manière d'incarner deux tendances épistémologiques représentatives de courants disciplinaires. Les deux auteurs choisis ne doivent néanmoins pas être réduits à la description qui en est donnée dans les lignes qui suivent.

Je tenterai ensuite de répondre à ces deux critiques qu'il faut prendre au sérieux. Tout d'abord, l'attention portée à l'ancrage biologique ne nous condamne pas à nous contenter de ce seul niveau de description, bien au contraire. Ensuite, le projet de naturalisation de la rhétorique nous engage à une réflexion épistémologique sur le statut de la rhétorique comme discipline, mais aussi sur la question du découpage disciplinaire. La rhétorique est-elle une pratique, un art, une science ou une philosophie? Plus largement, quelle discipline est-elle en charge d'étudier les phénomènes rhétoriques? La philosophie? La linguistique? La psychologie ou les neurosciences? Plus largement encore, se définit-elle par son objet? (Le discours public), par sa méthode? (L'analyse des discours) ou par un problème? (Quelles sont les conditions de l'efficacité de la persuasion?).

Cette série de questions me conduira à un développement spécifique sur le statut de la raison pratique au sein de la culture de l'artisanat dans laquelle est née la rhétorique. Je terminerai mon propos par une description de la persuasion dans le cadre de la réflexion générale que j'aurai menée.

Mon principal argument est que cette description de la rhétorique engage à un élargissement du concept de rationalité qui doit tenir compte davantage de la dimension pratique de l'intelligence humaine. Pour ce faire, je m'aiderai de certaines avancées des sciences cognitives, en les reliant à la description aristotélicienne, elle aussi naturaliste. Enfin, je plaiderai pour une pratique sérieuse de l'interdisciplinarité qui offre, à mes yeux, une voie nécessaire au développement humaniste de la recherche en sciences de l'homme.<sup>1</sup> En somme, cette habitude de penser l'homme dans toutes ses dimensions, laquelle exige du chercheur une démarche interdisciplinaire par défaut a tellement perdu de sa pratique qu'elle empêche parfois le chercheur de penser ensemble des phénomènes pourtant solidaires, mais désormais réputés relever de disciplines différentes. Je tenterai dans cette perspective de défendre l'idée qu'un retour à Aristote nous permet de nous réconcilier avec un homme complet dont la pratique rhétorique témoigne empiriquement que celle-ci relève de *toutes les disciplines*.<sup>2</sup>

- 1 Il faut noter encore que cette perte d'habitude de pratiquer une recherche humaniste dans laquelle tout chercheur se doit d'être un connaisseur plus ou moins éclairé des différentes disciplines qui touchent à la compréhension de l'homme est tellement tombée en désuétude que le fait même mérite d'être observé. Pour un chercheur, sortir de sa discipline est toujours perçu comme plus ou moins suspect. La charge émotionnelle qui accompagne ces polémiques à caractère épistémologique en dit long, là aussi, sur l'état de notre culture disciplinaire. Quoi qu'il en soit, il est un courant important des sciences cognitives qui «redécouvre» un certain rapport aristotélicien aux phénomènes humains, un rapport qui ne craint pas de placer l'homme au centre de la réflexion sans pour autant le réduire à sa seule dimension biologique. A mes yeux, l'un des représentants les plus intéressants de ce courant actuel en France est le physiologiste Alain Berthoz. Voir, par exemple, Berthoz (2009).
- 2 Selon l'affirmation même d'Aristote, qui pourtant nous a livré trois traités consacrés uniquement à l'art oratoire. Il reste que ces trois traités débordent de considérations d'ordre psychologique, éthique, sociologique, historique mais aussi biologique, logique, etc. Ainsi, la question du découpage disciplinaire est déjà fortement présente à ce moment.

## 2. Le risque des «impostures intellectuelles»

Le problème qui m'intéresse peut s'introduire rapidement en rappelant le fameux canular dû à Alan Sokal. En 1996, ce physicien américain envoie à la revue postmoderne *Social Text* un article qui parodie le style de certains intellectuels «encore» en vogue à l'époque. En particulier, l'article de Sokal utilise à foison le langage physico-mathématique pour «formaliser» des concepts de sciences humaines. Une fois l'article accepté dans la prestigieuse revue, le canular est dévoilé et le scandale éclate. Les intellectuels incriminés réagissent, la polémique s'emballe.

En première analyse, il s'avère intéressant de dresser une liste succincte des critiques que s'adressent les deux camps.

Au post-modernisme, on reproche principalement trois choses. Tout d'abord, l'usage immodéré et revendiqué de la métaphore (réputée floue) pour dénoter des phénomènes complexes de la réalité sociale. Ensuite, une sortie intempestive de l'enceinte disciplinaire jugée légère et inadéquate selon les canons de la science moderne. Finalement, une imposture généralisée qui prétend ériger le flou en critère de complexité et indice puissant d'intérêt intellectuel.

Dans l'autre camp,<sup>3</sup> on reproche au réductionnisme son matérialisme dépassé, sa naïveté face à la complexité du réel, son étroitesse d'esprit et finalement, reproche peut-être le plus fondé, le risque permanent de sombrer dans un «paralogisme naturaliste» qui cacherait mal ses relents idéologiques (la nature est bonne parce que c'est la nature). À partir de cette polémique et au vu de l'échange d'arguments auquel elle a donné lieu, il faut dès à présent faire une observation générale.

En effet, derrière une controverse qui aurait pu s'avérer extrêmement fertile pour la réflexion épistémologique, s'est manifestée une haine entre ces deux camps disciplinaires dont la virulence me paraît en soi préoccupante. Plus grave, la violence de l'échange semble avoir eu pour conséquence malheureuse pour la vie intellectuelle qu'elle a encore accentué les replis

3 Sans doute en partie parce qu'il s'agit d'une réponse à la critique formulée, on use et abuse d'un style pamphlétaire hérité du 19<sup>e</sup> siècle, où à peu près tous les coups sont permis, et cela, surtout s'ils portent en dessous de la ceinture.

disciplinaires dont la pensée humaniste n'est pas sortie grandie. Quoi qu'il en soit, l'imposture qu'entendait dénoncer Sokal véhiculait une vision implicite de la rhétorique qui va maintenant nous occuper.

### 3. Le style postmoderne: Lacan et la rhétorique

Le psychanalyste Jacques Lacan est l'un des auteurs visés par le canular de Sokal. Sa position est intéressante à critiquer pour deux raisons. Tout d'abord, elle est représentative du courant structuraliste formaliste et parfois abstrait dont l'épistémologie s'accorde mal à un naturalisme hérité d'Aristote, lequel place spontanément l'homme concret au centre de la réflexion. Ensuite, la position de Lacan a représenté, si pas initié, le fameux *rhetorical turn* qui a contribué à colorer la rhétorique d'une épistémologie constructiviste très éloignée de sa version antique.<sup>4</sup>

Pour plus de précision, je me concentrerai sur un article d'Angèle Kremer-Marietti (2004) qui rapporte une discussion entre Chaïm Perelman et Jacques Lacan à propos de la métaphore (1961): C'est surtout son commentaire qui retiendra mon attention: «Je ne me sens aucune compétence à expliquer la formule (algébrique?) posée ensuite par Jacques Lacan, mais, me contentant de ce qu'il en écrit [...]» (Kremer-Marietti 2004: note 27).<sup>5</sup>

Par-delà l'ironie de son commentaire, on comprend surtout que la «formule» proposée par Lacan est, aux yeux de Kremer-Marietti, aussi opaque qu'inapte à formaliser quoi que ce soit. Il est cependant important de ne pas passer à côté d'une motivation présente chez de nombreux structuralistes, en particulier, lorsqu'ils s'occupent de psychologie ou de

4 Selon l'historien Carlo Ginzburg, la lecture constructiviste de la rhétorique commence déjà chez Nietzsche. Voir Ginzburg (2002). Ce n'est pas le lieu d'entrer dans cet aspect du débat pourtant passionnant. Retenons néanmoins l'ouvrage magistral de Cassin (1995), qui a été un élément important dans le développement d'une épistémologie constructiviste associée à la rhétorique. Pour une critique admirative du texte de la grande philologue, je me permets de renvoyer à Danblon (2002).

5 Pour le lecteur courageux, voici la fameuse formule lacanienne:  $S/S^1.S^2/X \rightarrow S(1/s^m)$ .

psychanalyse. Une motivation décrite avec la plus grande clarté par la philosophe à propos de la conception lacanienne de la rhétorique:

Car ce qui est assez remarquable et rarement (ou jamais) avoué, c'est que Jacques Lacan a tenté avec beaucoup d'autres philosophes et écrivains du 20<sup>e</sup> siècle d'éluder (et sans y réussir mieux que les autres) la question du sujet et, à la limite, il a même dû souhaiter effacer en psychanalyse tout recours légitime à la subjectivité. D'un commun accord dans lequel sans doute cette génération de structuralistes voyait une planche de salut pour la philosophie du 20<sup>e</sup> siècle. (Kremer-Marietti 2004: 126)

Ce commentaire me paraît très juste, en même temps qu'il souligne l'immense paradoxe pour une discipline, de fonder sa scientificité sur la disparition du sujet, alors même que c'est lui qui se trouve au centre de la réflexion. Quoi qu'il en soit, le *rhetorical turn* a en effet semblé constituer pour une génération de philosophes une «planche de salut» pour une discipline qui devait donner des gages d'objectivité et de scientificité. En même temps, la motivation éthique ou psychologique d'éviter cette fois une dérive idéologique de sinistre mémoire, se trouvait assouvie par la disparition de la question la plus délicate, mais sans doute la plus intéressante à traiter: le sujet humain lui-même.

Dans cette perspective, le langage mathématique ou physique, en ce qu'il n'est *pas* le langage *naturel*, devient un candidat idéal pour élaborer de nouvelles théories plus *objectives* sur l'homme. Le langage formel, en principe peu propice à décrire l'humain, incarne pourtant, aux yeux de la raison moderne, le *summum* de la science. C'est d'ailleurs ce dernier argument qui fut retenu par les dénonciateurs d'impostures intellectuelles. Pourtant, ce «refoulement» du sujet hors de la raison humaine, opéré par le structuralisme, doit nous aider à comprendre une partie de l'actualité des débats sur la question, même si les «sciences de l'homme» ont évolué de façon spectaculaire ces dernières décennies, en particulier grâce aux progrès des sciences cognitives.

#### 4. Le style réductionniste: Sperber et la «rhétorique»

Le structuralisme commençant à passer de mode, de nombreux auteurs, philosophes, anthropologues, psychologues ou linguistes ont pris le parti de se poser la question de la raison humaine en profitant du tournant cognitiviste et de sa variante évolutionniste. Dans cette perspective, on pouvait envisager le retour du sujet au sein d'une option naturaliste sans que cela n'implique nécessairement de dérive idéologique. Après Grice (1989) et son *Principe de Coopération* auquel on a pu reprocher un paralogisme naturaliste idéalement pacifié par des maximes kantienne, il fallait pouvoir assumer l'option naturaliste en évitant les accusations d'angélisme ou de rousseauisme. C'est l'option qui semble avoir été choisie par Dan Sperber pour la conception qu'il donne de l'argumentation, dans un article qui expose clairement sa position. L'auteur s'attache à une hypothèse qui se limite à l'argumentation (et au témoignage que je ne traite pas ici) sans prétendre englober l'ensemble de l'activité rhétorique:

Ma thèse principale sera que de nombreuses croyances socialement acquises sont fausses à cause non pas d'un mauvais fonctionnement, mais d'un bon fonctionnement de la communication sociale. Je soutiendrai en particulier que la possibilité de manipuler cognitivement autrui est l'un des effets qui rendent le témoignage et l'argumentation adaptatifs. (2007: 105-106)

Ainsi, dans l'approche évolutionniste de Sperber, on observe un retour du naturalisme (dont témoigne la théorie de l'évolution) en même temps qu'une prise en compte de la question du sujet humain (dont on s'attache à décrire les motivations à agir). L'adaptation, comme composante de la théorie de l'évolution, devient ainsi un critère pour guider la réflexion sur les mécanismes de la persuasion. Mais ici, le retour du sujet se fait par la voie de la manipulation. Si l'argumentation, qui produit la persuasion, a été sélectionnée, c'est parce qu'il est toujours plus efficace, dans une interaction, de manipuler autrui. Certes, Sperber ne dit pas que la manipulation constitue la visée de l'argumentation, ce qui serait absurde. Mais il ne nie pas l'idée que cette fonction aurait pu être favorisée pour sa haute valeur adaptative. Cette option résonne comme le miroir inversé du Principe de Coopération gricéen. Le retour du sujet se fait ici, non pas à travers l'image d'un idéal de coopération, mais plutôt à travers une sorte de *factualité* du mal dont le

corollaire éthique a pour nom *banalité*. Certes Sperber ne prend pas position sur la question. Ce n'est pas son propos. Il en fait le constat «objectif», c'est-à-dire «scientifique». Dans ce court article, d'ailleurs, l'étayage argumentatif est sommaire.<sup>6</sup> Mais voici certainement le plus intéressant dans les conclusions de Sperber:

À l'aune d'une norme logique ou épistémologique, le sophisme est un usage pervers de l'argumentation, une pratique qui en contredit la raison d'être. Dans la perspective évolutionniste esquissée ici, le sophisme est une manière malhonnête d'utiliser la «stratégie de monstration honnête» pour le grand bénéfice du locuteur. Autrement dit, la sophistique contribue à rendre l'argumentation adaptative. (2007: 105-106)

On le voit, Sperber ne saurait être accusé de colporter un stupide paralogisme naturaliste. Au contraire, en un mouvement, il classe la sophistique comme l'incarnation de la manipulation et affirme que, contrairement à ce que certains ont eu la naïveté de croire, la norme, si elle n'est pas directement la tromperie, n'est en tout cas pas la coopération. Soucieux d'éviter le paralogisme naturaliste, l'objectivité scientifique de Sperber peut orienter le lecteur vers une lecture spontanément utilitariste – pour ne pas dire cynique – des rapports sociaux. Ainsi, dans ce retour du sujet que, désormais, l'on peut traiter «objectivement», la question éthique semble suspendue au point d'apparaître non pertinente dans le développement théorique. Comme le commente Daniel Heller-Roazen:

Notre monde est désormais l'image inversée de celui de l'Antiquité. Ceux qui s'égarant du côté de l'excès de sensation 'ne se rencontrent pas vraiment', et les individus qui, dans la vie, ne ressentent pas la douleur ou la joie 'qui convient' se font plus nombreux chaque jour. [...] Une insensibilité commune, qu'Aristote aurait jugée bien 'inhumaine' est devenue la règle [...]. (2011: 320)

Ce lien établi entre sensibilité et sensation, d'une part, et absence de considération éthique dans le comportement, d'autre part, n'est en effet pas directement traité chez Sperber, comme si, en un sens, la question dépassait du cadre de la science. Les relations entre sensation et éthique sont pourtant longuement explicitées par Heller-Roazen, qui les replace dans le modèle

6 Il arrive parfois que ces «faits» scientifiques sur la banalité du mal ou sur la manipulation soient étayés par des expériences psychologiques ou par des statistiques réalisées à partir de témoignage. Peu importe ici la méthode utilisée pour «garantir» la scientificité de l'hypothèse.

aristotélien de la raison humaine. L'auteur nous invite d'ailleurs à prendre la mesure du fossé qui s'est creusé entre la culture moderne et la pensée antique sur ces questions qui touchent aux fondements de la raison humaine. Retisser ce lien perdu devrait nous permettre d'indiquer une troisième voie pour la raison humaine et, partant, pour le rôle central qu'y joue la rhétorique. C'est ce que je me propose de développer dans la suite.

## 5. Est-il rationnel d'être *touché*? Remarques sur le naturalisme d'Aristote

Il faut tout d'abord rappeler que la pensée moderne fut particulièrement anti-rhétorique. Sans doute, la réflexion contemporaine sur la raison, le raisonnement, l'argumentation et la persuasion subit encore de plein fouet les conséquences de l'absence de la rhétorique au cœur de la Modernité. En témoignent les nombreuses publications qui s'intéressent aux liens entre parole publique, raison et action, issues de toutes disciplines qui bien souvent ignorent totalement la tradition rhétorique, en particulier, dans sa version humaniste héritée d'Aristote. A ce titre, l'étude de Daniel Heller-Roazen sur l'histoire des liens entre raisons et sensations, si elle n'utilise pas davantage la rhétorique que ses contemporains, nous fournit pourtant un angle intéressant pour repenser les liens entre raison et persuasion, en déplaçant le curseur traditionnel et, pour tout dire, en l'éloignant de la question lancinante de la manipulation. L'auteur rappelle ainsi *a contrario* comment la coupure entre sensations et entendement, lors du passage des modèles antiques aux modèles modernes, a tracé une frontière dont nous sommes toujours les héritiers. Plus précisément, il affirme qu'au *cogito* cartésien, généralement admis comme fondement de la raison humaine, s'oppose une version aristotélienne, que l'on peut énoncer selon la formule: «Je sens, donc je suis».<sup>7</sup> L'auteur fait ainsi valoir que dans la pensée aristotélienne, le fondement de la raison humaine – la raison de l'homme, en tant qu'animal

7 Chapitre 5: «*Sentio ergo sum*», en particulier, p. 63.



politique – est un «sens commun» (*sumaisthèsis*). Un sens commun, à la base de l'éthique, qui, pour le dire rapidement, serait l'analogie antique de la très intellectuelle «conscience» des Modernes (Heller-Roazen 2011).<sup>8</sup> Or ce sens commun, évidemment, s'exerce par les sens (les cinq sens, mais aussi notre capacité à les faire interagir et à acquérir un regard réflexif sur cette interaction). En somme, de cette expérience tellement humaine, nous finissons par nous sentir humains par le fait même d'être *touchés* de la réalité de ce «sens commun». Sans doute, ce retour de la conscience dans le corps,<sup>9</sup> avec sa dimension éthique, semble propice à éviter les deux écueils qui ont été évoqués. L'on pourra ainsi éviter la solution postmoderne qui voit dans la persuasion des «effets» du discours apparemment déconnectés des sujets humains. Mais l'on pourra dans le même temps dépasser un cognitivisme trop étroit qui affirme un lien supposé naturel entre l'efficacité de la persuasion et les contenus épistémiques véhiculés. Dans sa version optimiste, l'efficacité témoignerait de la vérité, dans sa version pessimiste, elle serait intimement liée au mensonge et à la tromperie. La très ancienne polémique touchant aux liens entre persuasion et manipulation, une polémique dont les bases n'ont guère changé depuis Platon, se décline aujourd'hui dans les termes d'une nouvelle forme de réductionnisme qui évite difficilement l'écueil du parallogisme naturaliste. Je crois pourtant que la voie tracée par Aristote, dans une certaine mesure, autorise un naturalisme non réductionniste.<sup>10</sup>

8 Cette inversion de la perspective sur la conscience fait d'ailleurs l'objet d'une grande popularité dans les sciences cognitives, en particulier depuis les travaux d'Antonio Damasio qui résonnent aussi comme une «mise au point» sur l'histoire de l'épistémologie de la raison humaine. Je pense bien sûr aux titres évocateurs: «L'erreur de Descartes», «Spinoza avait raison».

9 Un retour que l'on doit sans doute à l'impulsion des sciences cognitives après tant de siècles dualistes. Un retour qu'appelle d'ailleurs déjà de ses vœux Chaïm Perelman dès l'introduction à sa *Nouvelle Rhétorique* (voir Perelman & Olbrechts-Tyteca 2000).

10 Il faut être prudent dans cette affirmation. Comme tous les grands auteurs, il arrive qu'Aristote soit au moins ambivalent, si pas contradictoire. Or, le paradoxe est à son comble lorsqu'on considère que c'est précisément l'auteur de la *Rhétorique* qui échappe le moins à un certain réductionnisme épistémologique, quand l'Aristote du *De Anima* ou de l'*Éthique à Nicomaque* développe avec finesse une vision de l'homme qui marie avec succès le naturalisme et l'humanisme. En somme, ma proposition consiste à inviter cet auteur à dialoguer avec son «double» plus rigide, auteur de la *Rhétorique*.

Quoi qu'il en soit, la leçon des Grecs tient au fait qu'ils nous incitent à repenser ensemble des états ou des activités apparemment aussi divers que *sentir, contempler, découvrir, persuader, être persuadé, délibérer* et enfin *agir*...<sup>11</sup> Il s'agit de penser ces activités humaines sans les placer d'emblée de part et d'autre d'une frontière qui définit depuis si longtemps notre épistémologie dualiste.

## 6. Retour sur la dichotomie *persuader* vs. *convaincre*

Ainsi, le déplacement du curseur à propos de la limite entre persuasion et manipulation consiste à focaliser la réflexion sur la dimension pratique de la raison humaine, ce qui nous conduit à présent à revenir sur la plus célèbre des oppositions en rhétorique: celle qui prétend discriminer entre la persuasion et la conviction.

Voici ce qu'en disent Chaïm Perelman et Lucie Olbrechts-Tyteca lorsqu'ils ont eux-mêmes redécouvert cette tension: «Une distinction classique oppose les moyens de convaincre aux moyens de persuader, les premiers étant conçus comme rationnels, les seconds comme irrationnels, les uns s'adressant à l'entendement, les autres à la volonté» (1989: 66).

Ainsi, selon cette dichotomie classique, ce qui relève de la volonté s'adresserait à l'irrationnel. Mais Perelman et Olbrechts-Tyteca s'engagent ensuite dans une discussion qui explicite la différence entre deux traditions dont la fameuse dichotomie est issue. Il y a tout d'abord les traditions philosophiques qui mettent l'accent sur le résultat: c'est typiquement le cas du pragmatisme à l'anglo-saxonne. Il y a ensuite, celles qui mettent l'accent sur les moyens: ce sont les traditions continentales et françaises. Chez les premiers, la persuasion est valorisée. Chez les seconds, c'est la conviction qui prend le dessus. Après en avoir explicité les contours, les auteurs proposent ensuite de mettre cette dichotomie en question. La suite de leur réflexion, exposée sous la forme d'un récit, présente un grand intérêt pour notre discussion. En effet, c'est en cherchant à dépasser le caractère trop

11 Pour une lecture très fine de ce vocabulaire grec et de ce à quoi il renvoie, on lira Klimis 2003.

contraignant de l'opposition entre «persuader» et «convaincre» et par l'observation empirique de la réalité des discours et des débats, que Perelman et Olbrechts-Tyteca affirment avoir redécouvert la rhétorique:

Ayant entrepris cette analyse de l'argumentation dans un certain nombre d'ouvrages, spécialement philosophiques, et dans certains discours de nos contemporains, nous nous sommes rendu compte, en cours de travail, que les procédés que nous retrouvions étaient, en grande partie, ceux de la *Rhétorique* d'Aristote; en tout cas, les préoccupations de ce dernier s'y rapprochaient étrangement des nôtres. Ce fut pour nous à la fois une surprise et une révélation. Perelman & Olbrechts-Tyteca 1989: 71.

Perelman et Olbrechts-Tyteca évoquent à la fois une surprise et une révélation. Cet aveu est suivi d'un commentaire sur les grandeurs et misères de la rhétorique au cours des siècles qui ont précédé leur propre réflexion et qui, bien sûr, ont nourri leurs propres *a priori*. Les auteurs voient en particulier un préjugé tenace dans le modèle des pseudo-dialogues de Platon:

L'illusion que produit cette méthode consiste en ce que, du fait que l'interlocuteur admet chaque chaînon de l'argumentation, on croit ne plus être dans le domaine de l'opinion, mais dans celui de la vérité, et l'on est convaincu que les propositions que l'on avance sont bien plus solidement fondées que dans l'argumentation rhétorique ou l'épreuve de chaque argument ne peut se faire. (Perelman & Olbrechts-Tyteca 1989: 82)<sup>12</sup>

Ce jugement posé par les auteurs du *Traité de l'argumentation* évoque le commentaire de Dan Sperber (2007: 105) à propos d'Alvin Goldman. Ce dernier, affirme Sperber, a lui aussi dénoncé ce modèle monologique comme idéal d'un raisonnement qui serait garanti par l'autorité d'une inférence produite en solitaire. En effet, la technique vantée par Platon et inspirée par son maître Socrate ne fournirait, nous disent les auteurs, qu'une *illusion* de vérité. Or celle-ci est liée au préjugé selon lequel le raisonnement serait idéalement monologique (et donc monotone au plan logique). À l'inverse, un modèle plus rhétorique nous incite à voir dans le raisonnement une action collective et idéalement dialogique, dont le processus, loin d'être monotone, est semé d'essais et d'erreurs, d'une part non négligeable de navigation à vue, d'hésitations et de réajustements. À l'évidence, dans un tel modèle, les erreurs et les impasses sont aussi rationnelles que les réussites précisément, en ce

12 On retrouve une part de cette illusion dans le modèle habermassien de l'argumentation, paradoxalement, alors que ce dernier insiste pourtant sur l'importance de la participation de tous au débat.

qu'elles conservent, au sein même du raisonnement, la mémoire (et donc la richesse) des inférences vécues et de la diversité des points de vue qui se sont affrontés. Les décisions prises de cette façon se révèlent ainsi plus éclairées, au sens précis où elles sont éclairées par les arguments d'Autrui.

Ainsi, Perelman et Olbrechts-Tyteca redécouvrent la rhétorique via une nécessité rationnelle d'élargir, mais aussi d'assouplir la notion même de rationalité. Or leur modèle, pourtant d'inspiration juridique, s'inspire parfois lui-même d'une conception «continentale» de la raison qui valorise la conviction sans trop savoir que faire de la persuasion, si ce n'est un parent pauvre de cette même raison. Je crois pourtant qu'il faut aller un pas plus loin vers la pensée aristotélicienne pour enrichir la réflexion contemporaine sur la persuasion.

## 7. Persuader et être persuadé en retour

Aristote lui-même entame sa *Rhétorique* en rapprochant tout d'abord cette *technè* du modèle de la dialectique et donc d'une conception logique et syllogistique du raisonnement (*Rhétorique* I, 1, 1354a.). Plus loin dans le traité, on trouve pourtant des définitions de la rhétorique qui placent la persuasion au centre de l'activité. Il y est question, là d'une dimension heuristique (*ibid.*, I, 2, 1356a), ici d'une dimension actionnelle et même agonistique (*ibid.*, I, 1, 1355b).

En effet, dans le système aristotélicien, la persuasion s'inscrit dans l'exercice d'une action pratique qui dépasse de loin l'usage de la parole. C'est le produit de cet exercice qui doit déboucher sur une disposition à la décision et plus largement, sur une disposition à l'action citoyenne. Dans cette optique, une compétence se développe qui présente deux aspects: il est aussi important de persuader que d'être persuadé en retour, puisque les rôles d'orateur et d'auditoire, liés à la contingence de la vie citoyenne, sont constamment échangés.

Je crois qu'on ne comprend la nature de la persuasion, pierre angulaire des débats sur la place publique, que par ce détour important. Un détour qui nous conduit à rassembler au sein du système aristotélicien, bien davantage

que la *Rhétorique* et la *Dialectique*. Il y a aussi l'éthique, le sens commun, la raison pratique et l'action en général. Le cadre qui permet de penser la persuasion autrement que comme la production d'effets du discours ou comme une manipulation ou une tromperie est celui dans lequel est née cette technique. Ce cadre est celui de l'artisanat.

## 8. La persuasion dans les trois niveaux de l'art rhétorique

Pour comprendre ce cadre, il faut revenir à la conception grecque de l'art et à sa culture, à partir de laquelle l'a pensée Aristote. Ce retour – qui n'est pas un détour –<sup>13</sup> nous permettra d'élargir une conception moderne de la raison tout en explicitant les liens entre les différents domaines de la cognition humaine où intervient la persuasion. Dans cette conception «technique» de l'art, trois niveaux distincts se dégagent pour la raison pratique.<sup>14</sup>

Il y a d'abord l'ensemble des règles à suivre. Il y a ensuite la production, la visée. Il y a, enfin, la disposition d'esprit que développe celui qui l'exerce. Voici comment peuvent se décrire ces trois niveaux pour la rhétorique:

<b>Technè (art)</b>	<b>Praxis (action)</b>	<b>Poièsis (production)</b>	<b>Hexis (disposition)</b>
<b>Rhétorique</b>	Construire les preuves dans le discours	Persuader	Savoir persuader et être persuadé en retour

Table 1. Les trois niveaux techniques de la rhétorique

- 13 Je veux dire par là que ce qui peut apparaître comme une sortie intempestive des remparts disciplinaires me paraît une démarche nécessaire pour étayer l'argument que j'essaye de développer concernant la spécificité de la raison pratique. À mes yeux, la culture de l'artisanat n'est pas un cadre métaphorique pour penser la question. Il s'agit d'un cadre anthropologique qui devrait s'avérer pertinent pour penser toutes les activités pratiques de la raison humaine.
- 14 Sur tout ceci, on lira Vernant (1965), en particulier, le chapitre 4: «Le travail et la pensée technique», pp. 263-322 et Sennett (2010).

Ce tableau permet tout d'abord d'exposer les trois niveaux de description qui vont se révéler pertinents pour une bonne compréhension de la raison pratique. Dans ce dispositif, l'auditoire n'est pas un réceptacle passif du discours, mais un partenaire qui interagit au sein même de l'activité et qui, à ce titre, participe des différents niveaux qui vont être décrits.

### 8.1. *La praxis: des actions intelligentes coordonnées*

Commençons par le premier niveau, celui de la *praxis*. Il s'agit de décrire l'aspect le plus pratique, celui de l'ensemble des activités concrètes qui peuvent toutes s'énoncer sous forme de règles à suivre. Mais ces règles ne s'appliquent pas pour autant de façon mécanique. L'orateur construira lui-même les preuves de son art en fonction de la situation concrète à laquelle il est confronté (situation déterminée, selon Aristote, par le genre, c'est-à-dire par l'institution dans laquelle la parole publique se produira). Dans le modèle de rhétorique démocratique qui l'a vu naître, le postulat de symétrie entre les citoyens est structurel. En effet, le débat public prévoit que les rôles d'orateur et d'auditoire se distribuent en fonction de la situation d'énonciation: chacun, en principe, possède une expertise équivalente dans l'usage des preuves, même si les rôles s'alternent au sein de l'activité. En outre, l'auditoire est actif *en tant qu'*auditoire puisqu'à l'issue de la délibération, c'est lui qui *décide*.

En somme, la *praxis* peut se décrire par la combinaison d'une compétence spécifique qui comporte ses règles propres. Celles-ci peuvent être énoncées au sein de traités de rhétorique, mais à côté des règles, il y a toujours l'expérience et l'intuition. C'est le volet non formalisable de l'art, dont l'artisan expérimenté a la maîtrise et qui débouche toujours sur des actions concrètes. Au-delà de la vérité et du mensonge, la question réellement pertinente pour ces techniques est celle, pratique, de l'efficacité. Cette question concerne le niveau de la *poiësis*, que je vais aborder maintenant.

### 8.2. *La poiësis: l'effet et la nécessité du détour*

Dans l'art grec, la *poiësis* désigne ce que l'art vise à produire. La persuasion comme visée est donc un état plus ou moins stable, même s'il ne peut être

envisagé comme définitif: il est toujours à reconquérir. En somme, dans ces matières, rien n'est jamais gagné. Cet aspect des choses qui peut paraître anecdotique permet pourtant de souligner le fait que tous les arts s'appliquent à un monde complexe, en principe, toujours ouvert.

C'est sans doute pour cela que la *poièsis* ne peut s'obtenir que de façon indirecte, comme un *effet* de l'application précise, patiente et rigoureuse de l'ensemble des pratiques offertes par la technique. Cette propriété de la *poièsis* permet de mettre en lumière l'un des aspects centraux de l'intelligence pratique qui se trouve en jeu. Une application mécanique de principes appris par cœur ne mène à rien. Toute technique demande une *attention* particulière, une concentration sur la tâche, qui pourrait être propice à induire la *poièsis* en tant qu'effet secondaire de la technique. On commence par imiter et, ce faisant, on acquiert peu à peu une maîtrise de l'action intelligente. Autrement dit, dans cette conception de l'art, il est rationnel de ne *pas s'acharner* à atteindre un but, c'est là une condition de l'efficacité de la visée poursuivie.

De son côté, Jon Elster mentionne des situations analogues qu'il décrit comme relevant des limites de la rationalité (Elster 1987). L'exemple classiquement cité évoque l'insomniaque qui «trompe» son insomnie en lisant de sorte qu'il obtiendra le sommeil par un effet secondaire: la lecture. Elster insiste encore sur le fait que cette activité *apparemment* irrationnelle est en fait très courante. A mon sens, elle souligne le caractère éminemment pragmatique de cet aspect de la raison humaine. Un aspect qui touche à la persuasion et qui ne peut être jugé irrationnel que dans une conception purement cognitive de la raison. Au contraire, la structure indirecte de l'action intelligente souligne l'efficacité pratique qu'il y a à se détourner de la visée poursuivie pour la rendre plus aisée à atteindre.

Finalement, de cette description du niveau *poïétique* de la persuasion se dégagent deux traits importants. Tout d'abord, il y a une action coordonnée qui peut se décrire en termes d'interaction entre deux protagonistes plutôt qu'en termes d'agent et de patient. Ensuite, la visée de la technique, sa *poièsis*, ce qui doit être produit, s'obtient en tant qu'effet secondaire et non pas en tant que but directement visé. La *poièsis* ainsi produite, la persuasion, est un état physico-psychique qui présente aussi un caractère dynamique. De ce point de vue, la persuasion dépasse le simple état mental à base de croyance, lequel enfermerait ultimement la description dans le couple vrai/faux. Ainsi, un dispositif indirect, qui paraît au premier abord peu intuitif, révèle son

caractère éminemment rationnel au plan de la raison pratique. Le critère, ici, est l'efficacité.<sup>15</sup>

### 8.3 L'hèxis: devenir un citoyen

Venons-en à présent au troisième niveau de description de la technique: celle de la création d'une disposition plus stable. Tout art au sens grec du terme contient trois niveaux pertinents qui décrivent l'exercice de l'intelligence pratique mise au service d'une technique: l'ensemble des actions intelligentes, la visée générale poursuivie, mais aussi une disposition plus stable qui devient, pour celui qui exerce régulièrement son art, comme une seconde nature. Mon hypothèse est que cette seconde nature recouvre simplement l'art d'être un citoyen, qui sait comment délibérer, parce qu'il sait comment persuader et être persuadé en retour, et parce que, pour cela, il aura tout d'abord appris par l'exercice de son *sens commun* à *toucher* et à *être touché* en retour.

Ces qualités qui deviennent une seconde nature rassemblent tout à la fois la théorie et la pratique, l'intelligence et l'éthique, la rigueur et la souplesse.

## 9. Conclusion

Finalement, c'est à la condition de cet élargissement de notre conception de la raison, à laquelle nous invite le modèle grec, que l'on peut songer à naturaliser la rhétorique, en évitant les écueils du post-modernisme et du réductionnisme. Mais il me reste pour conclure à revenir une fois encore à la question de la persuasion. La pratique de celle-ci sur la place publique développe-t-elle une seconde nature qui serait *en soi* manipulatrice? J'envisagerai pour ma part un point de vue différent de celui de Sperber, même si je pense qu'il convient de penser la raison humaine dans un cadre naturaliste, le seul qui rende *justice* à toute la complexité de la raison humaine.

15 Voir encore à ce sujet Berthoz (2009).



Il est utile pour cela de revenir à la figure du Sophiste. Comme le rappelle Jean-Pierre Vernant (1965), celui-ci se confond au départ avec la figure de l'artisan. Comme tout artisan, il développe cette forme d'intelligence pratique, cette agilité, que les Grecs nommaient la *mêtis*.<sup>16</sup> Pour tout artisanat concret, cette forme d'agilité est neutre au plan éthique, puisqu'il n'y a pas de sujet.

Mais dans les arts comme la rhétorique ou comme la médecine, la question éthique se pose nécessairement. Si la *Rhétorique* d'Aristote est déjà marquée par la méfiance d'une société vis-à-vis de la raison pratique, son *Ethique* insiste à foison sur la figure de l'homme prudent, le bon délibérateur, l'idéal citoyen: le *phronimos*.<sup>17</sup> Le prudent d'Aristote, en somme, est celui qui exerce sa raison pratique et qui, à force d'exercice, parvient à prendre, le plus souvent, des décisions éclairées. Il devient ce qu'il doit être pour vivre en société: un citoyen qui sait pratiquer son métier de citoyen. De ce point de vue, l'art de la persuasion n'est en soi ni manipulateur ni honnête. Il permet à tous de développer une activité citoyenne dans laquelle chacun décidera, dans le huis clos de sa conscience, s'il la place au service de la coopération ou de la manipulation.

## Bibliographie

- Aristote (1990): *Rhétorique*. Traduction et commentaires par Médéric Dufour, Les Belles Lettres, trois volumes, Paris.
- , (1990): *Ethique à Nicomaque*. Traduction et commentaires par Marcel Tricot, Vrin, Paris.
- Aubenque, P. (1963): *La prudence chez Aristote*. PUF, Paris.
- Berthoz, A. (2009): *La simplicité*. Odile Jacob, Paris.
- Cassin, B. (1995): *L'effet sophistique*. Gallimard, Paris.
- Château, J.-Y. (éd.) (1997): *La vérité pratique. Aristote, éthique à Nicomaque, Livre VI*. Vrin, Paris.
- Damasio, A. (1995): *L'erreur de Descartes. La raison des émotions*. Odile Jacob, Paris.
- , (2003): *Spinoza avait raison*. Odile Jacob, Paris.
- Danblon, E. (2002), *Rhétorique et rationalité. Essai sur l'émergence de la critique et de la persuasion*. Bruxelles, Editions de l'Université.

16 Voir plus précisément, Detienne & Vernant (1993).

17 Voir Aubenque (1967).

- , (2013): *L'homme rhétorique. Culture, raison, action*. Editions du Cerf, Paris.
- Detienne, M. & Vernant, J.-P. (1993): *La mètis des Grecs. Les ruses de l'intelligence*. Flammarion, Paris.
- Elster, J. (1987): *Le laboureur et ses enfants. Deux essais sur les limites de la rationalité*. Editions de Minuit, Paris.
- Ginzburg, C. (2004): *Rapports de force. Histoire, rhétorique, preuve*. Seuil, Paris.
- Goyet, F. (2009): *Les audaces de la prudence. Littérature et politique aux 16<sup>e</sup> et 17<sup>e</sup> siècles*. Garnier, Paris.
- Grice, P. H. (1989): *Studies in the way of words*. Harvard University Press, Cambridge/Londres.
- Heller-Roazen, D. (2011): *Une archéologie du toucher*. Seuil, Coll. La librairie du 21<sup>e</sup> siècle, Paris.
- Klimis, S. (2003): «Voir, regarder, contempler. Le plaisir de s'apprendre son humanité», in *La Poétique d'Aristote: lectures morales et politiques de la tragédie*. Etudes Philosophiques, PUF, Paris, 466-483.
- Kremer-Marietti, A. (2004): «Rhétorique sociale et métaphore du sujet. Perelman, Burke et Lacan», in Meyer, M. (éd.), *Perelman le renouveau de la rhétorique*. Collections «débat», PUF, Paris. Disponible en ligne à l'adresse:  
<http://www.psychanalyse.lu/articles/MariettiPerelmanLacan.htm#fn27>.
- Perelman, Ch. (1961): «L'idéal de rationalité et la règle de justice», *Bulletin de la Société Française de Philosophie*, 55<sup>e</sup> année, n°1, 29-33.
- Perelman, Ch. & Olbrechts-Tyteca, L. (1989) [1950], «Logique et rhétorique», in Perelman, Ch. (éd.), *Rhétoriques*. Editions de l'Université, Bruxelles, 6-107.
- Sennett, R. (2010): *Ce que sait la main. La culture de l'artisanat*. Albin Michel, Paris.
- Sperber, D. (2007): «Le témoignage et l'argumentation dans une perspective évolutionniste», in Bouvier, A. & Concin, B. (éds.), *L'épistémologie sociale. Une théorie sociale de la connaissance*. Editions de l'EHESS, Paris, 105-106.
- Vernant, J.-P. (1965): *Mythe et pensée chez les Grecs*. La découverte, Paris.

# A case for emotion awareness

Evgenia PAPAOUNI, Université Libre de Bruxelles

## 1. Introduction

Blinded by passion, jealousy or anger? Inspired by love for freedom, getting courage through cherishing a purpose, self-abnegation and sacrifice for others? Be politically indignant? Adapt in society, remember, pay attention, interpret, justify, take a decision and reason? According to recent research findings (Damasio 1995, Blanc 2006, De Houwer & Hermans 2010), all these involve some form of emotion.

Albeit pervasive, emotion is controversial for argumentation studies as some theories of argumentation account of it as an obstacle to reasoning or as an instrument of ‘manipulation’.

The ultimate purpose of this contribution is to deal with both these claims on emotion from the point of view of a rhetorician. To do so, I start by presenting the divide over emotion in argumentation studies (section 2). Although the term ‘manipulation’ seems to be potentially argumentative in the eyes of a rhetorician, I try to come to terms with the *claim* of manipulation, taken to mean the unsatisfactory results of a debate, portrayed as deceptiveness by the participants in it (section 3). I focus in particular on the efficiency of emotion. Then I make a bibliographic review of contemporary research in psychology and theory of mind concerning the relation between emotion and cognition, to see if the divide existing in argumentation studies also exists among different schools of thought in experimental psychology and in theory of mind. Even studies which contrast logicality to emotion, in the 80s do not allow us to conclude that emotion is fallacious (section 4.3), as it sometimes appears to improve the agent's performance in the experimental task. From that, I draw attention to a contemporary paradigm shift in cognitive science, as far as the study of emotions and the perception of human rationality is concerned. According to the new paradigm, inaugurated by the seminal works of Damasio (1995),

rationality is composite and it includes emotions, which confirms ancient rhetoric's intuition (section 4.4).

As verbal interactions allow for conflict by definition, attempts to subvert a composite rationality can be accounted for in the new paradigm. Emotions can be involved in such attempts/ unsatisfactory outcomes, as they involve a rapid appraisal of a situation and do not always avail themselves of propositional content (section 4.5).

On the basis of such an extended view of rationality, I propose an alternative way out of the 'manipulation' claim conundrum, through debate on emotions themselves. In such a debate emotion is not necessarily contrasted to logicity, but to another emotion, more suitable (section 5). Rhetoric can remain faithful to its descriptive vocation and still avail itself of a pedagogical project. It is through familiarisation with rhetoric, theory and practice, as well as through awareness of our own emotions that we can become more satisfied with our communication practice (5.1). It is not a matter for the rhetorician to whistle blow manipulation. Claims of manipulation are tested in the communication practice itself (5.2).

## 2. Emotion and argumentation theories

Despite the contemporary revival of the interest in emotion in various fields of the natural and social sciences, no unanimity exists, in argumentation studies, on the legitimacy and treatment of emotion. From this point of view, we could schematically classify argumentation theories as follows: (a) theories that are not interested in emotion, as they exclusively study the logical structure of arguments; (e.g. S. Toulmin 1958, J. Woods 2004) (b) theories that consider *pathos* as a means of proof, like classical rhetoric and most new rhetoricians (e.g. Plantin 2000, 2012); (c) theories that see the appeal to emotion as a fallacy or as strategic manoeuvring (e.g. Van Eemeren & Grootendorst 1994; 2006); (d) theories that aim at evaluating emotions' legitimacy under specific conditions (e.g. Walton 1992).

Aristotle qualifies *pathos* as one of the three legitimate discursive proofs and gives ample instructions on its usage in his third book of Rhetoric.

Quintilian more strongly recommends the use of *pathos* for the sake of efficiency. Contemporary rhetoric studies meet again with the classical tradition.<sup>1</sup>

In the fifties, Neo-Rhetoricians tended to neglect *pathos*.<sup>2</sup> Nowadays – and probably in parallel with the changing paradigm in all social sciences – rhetoric turns to the Aristotelian *pathos* in order to account for the role played by the speaker's and the interlocutor's emotions in persuading an audience and modifying an attitude. Emotions are not only a means of proof, but also a matter of debate in itself, since their individual appropriateness to the situation is questioned by participants in argumentation (Plantin 1998, Micheli 2010). *Pathos* is a central analytical category that translates speakers' subjectivity in situated speech.

In contrast to Neo-Rhetoricians, normative approaches to argumentation adopt a suspicious stance towards emotion, a stance that ranges from total rejection to acceptance under conditions. First-wave pragma-dialectics considers appeals to emotion to be fallacious (van Eemeren & Grootendorst 1994). The extended theory allows for strategic manoeuvring provided the rules of critical discussion are complied with. Persuasion is tolerated as long as it does not take priority over critical discussion (van Eemeren & Houtlosser 2006). We owe the most systematic normative approach to emotion to Walton (1992: 27-28). Walton focuses specifically on emotion trying to give it some legitimacy through a comprehensive theory rather than dealing with it as an obstacle or a side product (and necessary evil) in argumentation. He proposes a case by case evaluation of emotion in argument along a double pragmatic criterion: its effects on the argumentative goal of the participants and the relevance to the debated matter. If the goal of argumentation is promoted, through revealing deep commitments and through motivation for action, appeals to emotions are legitimate and beneficial. If the dialectical goal is diverted and *pathos* leads to suppression of evidence and entrenched positions, emotions are presumed to be fallacious.

This balanced approach has the merit of contextualising emotion and linking it to six variations of critical dialogue (1992: 19-22) which remotely

1 For a substantiated account of the historical positioning of argumentation theories on emotion, see Plantin (1998, 2000), Micheli (2010).

2 Perelman talks about values rather than about emotions.

remind us of ‘genre’. Nevertheless human communication offers a huge amount of ‘genres’ that cannot be reduced to the six variations of critical dialogue proposed by Walton.<sup>3</sup> More importantly, an approach that defines argumentation as problem/conflict resolution excludes broader definitions of argumentation that do not take scientific discourse as a model. This exclusion does not answer the question why people engage in argumentation over problems they know to be unsolvable or without entertaining any expectation to convince. People may indeed engage in argumentation to assert their position, to bond with others or to underline difference of opinion which has to be resolved through voting.<sup>4</sup> As a result, emotion in ‘genres’ outside the six variations is unnecessarily disqualified.

This schematic overview shows the division of argumentation studies over the role of emotion in argument. Before I turn to cognitive science for some support on the role of emotion in argument, I will touch upon another non-consensual topic in argumentation studies: Manipulation in terms of normative versus descriptive theories.

### 3. The claim of manipulation

Rhetoricians, in general, abstain from offering systematic norms to identify ‘manipulation’ in the name of a descriptive stance. They are even cautious towards the term manipulation as they consider it to be a metalinguistic argument, one pointing to the frame and not the content of the matter. C. Plantin underlines that one will never attribute the manipulator’s role to oneself (2002: 239). He considers the claim of fallacy to be an argument in itself too, that he ironically calls ‘*ad fallaciam*’ after the long series of fallacies historically identified under their Latin name (1995).

3 These variations are: critical discussion, information seeking dialogue, negotiation dialogue, inquiry dialogue, scientific inquiry, quarrel (*ibid.* p.20-21).

4 ‘Same sex marriage’, ‘taxing financial transactions’, ‘deregulation of labour market’ are topics on which the conflict is not reasonably expected to be resolved through argumentation. These topics are intertwined with values and interests.

Nevertheless, the derogatory term ‘manipulation’ exists in language and we are all familiar with claims of deceptiveness on behalf of an observer of the situation and of the ‘victim’ oneself with the benefit of hindsight. Some of these claims target speech that may involve the use of emotions. Take for instance the following examples:

- (a) This new car really matches your social status. (need to feel and seem superior);
- (b) Immigrants steal our jobs. (fear, feeling of unfairness);
- (c) We are the best. (pride).

All these phrases may be inaccurate or subjective. They can also elicit in the addressee in potentially desirable or undesirable behaviours, in his eyes and/or in the eyes of society. The same phrase may incite both a desirable and an undesirable behaviour: (c) does not lead to the same result if the ones who are the best are a soccer team or a nation on the verge of war. The car buyer may be absolutely satisfied with his choice, although in our eyes he spent a fortune for nothing. And whether we like it or not, Europeans are not unanimous on the number of immigrants a country can assimilate: We see that the addressee and an observer may not agree on whether the phrase has been deceptive.

Besides, when the claim of manipulation is made on behalf of an observer, we can infer the observer’s own (opposite) position; we can also infer (semantically) that the person who makes the claim (often a representative of a normative theory of argumentation) portrays the addressee of the manipulative phrase as a victim in need of protection, probably through some norms/discourse ethic on what is inappropriate in argumentative speech.<sup>5</sup> Moreover, we infer that if such a discourse ethics prevailed and its rules were respected, people would agree more often among them, they would not live beyond their means and they would not be xenophobic or imperialistic any more. Due to these semantic inferences from the term manipulation, I consider it to be argumentative in the sense of Anscombe & Ducrot (1983).

5 The typical example of manipulator in literature is the one of politicians that led their people in wars through demagogy. I consider this example to be politically argumentative as it absolves the citizens of the respective countries from their responsibility.

In a nutshell, I believe that the above statements (a) to (c) are not sufficient per se to prove that there is manipulation. They have to be interpreted in a specific context. This is why I am referring to claims of manipulation. I do this only to the extent that such claims are used by scholars and laymen to dismiss the explicit or implicit appeal to emotions in discourse.<sup>6</sup>Now, what is it that makes emotions so powerful and alarming?

The following questions are open:

- (i) If the use of emotion in argument is fallacious according to pragma-dialectics, why are audiences still swayed by it?<sup>7</sup>
- (ii) How does a self-poised descriptive/rhetoric theory of argumentation account for manipulative scenarios, i.e., situations where agents themselves recognise that they have been driven astray by skilful communicators? Does their proclaimed ‘descriptive’ stance amount to permissiveness? I will come back to the issues under (ii) in section 5 of this article.

Rhetoricians circumvent the questions under (i) by adopting a descriptive approach (not reasoning in terms of fallacies), by admitting *pathos* to be a legitimate means of proof and by referring to psychology for the study of the cognitive mechanisms behind persuasion. This approach though does not come to terms with an empirical phenomenon: the situation where audiences feel themselves either misled by skilful communicators through the use of/appeal to emotion or trapped by their own emotional predispositions, independently of their interlocutor’s intention, or both things at the same

- 6 Emotion can be explicit or implicit. Rhetoric discusses the emotion of the audience/eceiver under the category of *pathos* and the emotion of the speaker/source of the message, under the category of *ethos*. I don't make any differentiation between explicit or implicit emotion, present in the text, attributed or provoked, as these distinctions do not matter here.
- 7 The first question reminds us of a famous saying by Pareto (1968/1916) according to which “logic explores why a syllogism is erroneous while sociology is interested in the reasons why it is frequently accepted”. A. Bouvier (1995) points out that this remark boils down to a distinction between two different research programs, as far as fallacies are concerned: the exploration of their cognitive and their affective causes, respectively.



time.<sup>8</sup> In more general terms, we would like to test rhetoric's answer to the users' critics on the unsatisfactory outcome of a debate where emotions are involved.

Apart from answering such legitimate complaints, a second reason to deal with the claim of 'manipulation through emotion' is the need to test a theory's implicit model of human rationality. Theories of argumentation in the past were much too often influenced by the Cartesian dichotomy between the body and the mind, in the sense that they perceived reasoning to be emotionless and that they aimed at isolating speech from the speaker and the situation, to retrieve pure argument. This dichotomy has been reinforced in the era of informatics through the so called computer metaphor of the mind. Nowadays, the dichotomy is revisited and criticised (Damasio 1995, Blanc 2006).

On the other side, even if we take as a basis a broader rationality which includes emotions, we can still admit the possibility of a subversion of it, be it intentional or accidental, be it at the production or at the reception side of communication, especially as verbal interactions allow for conflict. The concession to the possibility of subversion of rationality may look strange after the misgivings I expressed on manipulation at the beginning of this section. We will see in section 5 how it is possible to come to terms with attacks to rationality with no need for the analyst to adopt a normative stance.

For these reasons, instead of circumventing the questions under (i) and (ii) I will visit reported findings of cognitive experimental psychology in association with insights coming from the philosophy of mind on the relation between emotion and cognition<sup>9</sup> to answer them.

- 8 I use this rudimentary example, a piece of self-criticism produced by the participant in argumentation him/herself, for, as I already said, I consider the claim that *somebody else* is manipulated to be potentially argumentative in the sense of Anscombe & Ducrot (1983).
- 9 Traditionally rhetoricians are concerned with the discursive manifestation of emotions and, in the name of a 'clean job', leave to psychology those matters that relate to speakers' and audience's representations and feelings. Nevertheless, coming to terms with manipulation implies a return to interaction and the perlocutionary effects of speech on the audience. These effects are not necessarily intentional, as the term 'manipulation' seems to imply: they may be accidental, unpredictable, and may vary according to the audience.

## 4. Emotion and Cognition: from reciprocal influence to consubstantiality

### 4.1. *Some key ideas*

In cognitive science, the return to subjectivity and emotion is a relatively recent but dominant trend; the current literature on emotion is bulky, controversial, non-conclusive but promising.<sup>10</sup> I will limit myself to concisely underlying points of convergence that are relevant to the purpose of this chapter:

(a) It is admitted today that emotions are plural and composite. They consist in an appraisal of the situation, a preparedness to act, and several somatic and representational responses, i.e., they involve a cognitive, a motivational, and a physiologic component, all three acting simultaneously, in reciprocal influence. Appraisal concerns in particular the relevance of the stimuli for the individual and to the task, as well as the possibility to cope and to control. We can describe emotions through variables such as intensity, (positive or negative) valence, duration, familiarity and frequency. This definition fits the rhetorical use of pathos: adapt the speech to what is relevant for the audience, namely strong preferences, positive or negative.

(b) “The brain as a network of simultaneous and interactive operations replaced the computer as the dominant metaphor of the mind” (Chapelle in Blanc 2006: 63). Cognition is incarnated. Operations are parallel and not sequential. Emotion is not ‘noise’ in the system but part of it, meaning that we cannot escape the presence of some form of it. It seems to me that

10 Historically, emotion, when not completely discarded as a lower somatic expression in the dualist paradigm, has been studied separately from cognition for a number of reasons (Blanc 2006). It has been considered as a reminiscence of our animal past, a signal of alert in front of change that may paradoxically undermine performance and render our behaviour dysfunctional if it persists in high amounts. Cognitive science, in spite of its critic of this behavioural model, has also failed to come to terms with emotions, as it developed on the basis of a computational model of mind and, in parallel, considered individual functions (i.e., perception, memory and understanding) as lower and higher levels of a unique and linear processing. In addition to that, the difficulty to define emotions has complicated further a long-standing dichotomy between emotion and cognition.

according to this new paradigm and contrary to the computer metaphor, emotionless entities are corpses, autists or machines. When I am sitting calmly in my office looking at the sky, I do have some form of emotion not necessarily intense or recognisable, but still inseparable of what we call thoughts. As a consequence, we cannot deal with emotion in terms of present/absent. We have to deal with it in a continuum of intensity and description/qualification. This element looks important to me to argue that the attempt to banish emotion from the argumentative practice is lost labour.

(c) Lower and upper cognition limits are blurred by the fact that all levels interact. Memory and attention are seen as selection of information according to a relevance criterion. This fact underlines the subjective input of evidence used for such high-level tasks as priority setting, reasoning and decision making. Points (b) and (c) are crucial for argumentation theories, as each of them presupposes a model of human rationality which does not in all cases include emotions.

(d) Despite the emphasis on the importance of stimuli appraisal for the causation of emotions, automatic/non-conscious mechanisms attract more scientific interest at the moment (Blanchette & Richards 2010).<sup>11</sup> Automaticity and complexity explain that agents may disclaim ownership and control over the effects of emotion. In my view this explains historically the awe and suspicion of scholars and laymen towards emotion. A firm belief may be equally or even more erroneous and harmful than an emotion, but contrary to what happens in the case of emotions, we have the impression to control our own beliefs. We also can avail of more time to form a belief or an opinion, while emotion is spontaneous, rapid and unpredictable. These traits make emotions' strengths and weaknesses at the same time. Speed is important in appraising the relevance of a situation, but as we will see in section 4.5, appraisal is not knowledge.

(e) Individuals can provide verbal accounts of emotions and also (mostly narrative) scripts involving their causes and consequences (Blanc 2006), i.e., they have representations of their own, and other people's, emotions. Thus, what may start as an automatic/non conscious reaction can then be formulated into words, questioned and debated. This point is important for

11 This implies reconsidering both the effectiveness of experimental methods such as self-reporting and the rise of expectations from investigating techniques in neurobiology.

what I claim in section 5, namely that formulation into words and debate diminishes the risks associated with emotion.

(f) Emotions are a challenging test for scientific theories, terminology, models and research methods. Despite the role played by parallel, partly automatic and interactive operations in the causation of emotions (see points (b), (c) and (d)), research models, methods, parameters, models, and not least terminology, are still inspired by the old analytical paradigm that separates emotion from cognition.<sup>12</sup> For instance, the relationship between emotion and memory is studied in terms of linear causes and effects, in spite of significant evidence for consubstantiality. As a consequence, it is very difficult to talk about consubstantiality of emotion and cognition, even among scholars, as we are imprisoned in old analytical and linguistic categories.

#### 4.2. *Emotion and logicality in experimental psychology*

A good example of ambivalent relation between cognition and emotion (viewed as two separate phenomena) is the relation between memory and emotion. The effect of emotion on memory can be both distorting and enabling.

It has been experimentally shown that there is a link between emotional states and memory performance. According to these, emotion distorts, reconstructs and flaws memory. In general, emotional information attracts attention and is easily recalled. On the opposite side, a protective perceptual defence may be activated in the case of traumatic experiences. There is experimental evidence that past experience is recalled and interpreted according to present moods and attitudes (Bower *et al.* 1978, Mac Farland and Ross 1987, Spiro 1977, Uranowitz in Blanc 2006: 74, 91). Moreover, it is possible to induce false memories (Loftus 1978 in Blanc 2006). Cases have been reported of convicts who came to believe their own fake confessions (Blanc 2006: 96).

12 For instance, self-reporting is still used as an experimental method, although its efficiency to provide access to automatic processes has been questioned. In fact, it remains an important tool to access perceptions and the agents' own representations of emotions.

On the other hand, suppressing emotion worsens performance in memory tasks (Gross 1998 in Blanc 2006), creating a catch twenty-two for the seekers of emotion-free memory. A possible explanation for these observations is that emotion selectively orients attention to information. Subjects prefer inferentially coherent stories to fragmented information and, as we saw above, they avail themselves of narrative accounts/patterns when dealing with emotions.

Memory and attention influence higher order cognitive functions. There is evidence that interpretation of ambiguous stimuli is conditioned by the person's affective state, with anxiety-congruent interpretations being the most salient example (Blanchette and Richards 2010: 278). Empirical studies have also documented that affects influence risk perception (Johnson and Tversky 1983, Constans in Blanchette and Richards 2010).<sup>13</sup>

Emotion is shown to impair logic and deductive reasoning, probably because it diverts attention from the task when it is *incidental* to it. On the contrary, when emotion is inherent to the experimental task (*integral*), performance improves, perhaps due to the agent's better concentration. According to Blanchette and Richards (2010: 300-303) this paradoxical result has to be further studied. A rhetorician is sensitive to the role of emotion as a catalyst in performing a task as it looks obvious to him/her that emotion, as a link to the person's values and interests, plays a crucial role in persuasion.

Affective states are reported to have an influence over the strategy deployed in reasoning (Worth and Makie 1987 in Blanchette and Richards 2010). According to literature on 'depressive realism', sadness seems to be a predisposition for systematic processing (focus on the argument), while positive-valence emotions and anger favour heuristic processing (focusing on the source of the message and its superficial features). These last findings do not raise the question whether systematic processing is advantageous to heuristics and, if so, why.<sup>14</sup>

13 Anxiety congruent judgements probably appear at a higher extent when risk concerns the subject himself (Muris & van der Heiden 2006).

14 The association of emotion and cognition along the differentiation between central and peripheral information processing echoes the views of Chaiken (1980). Other models underline the role of motivation, involvement and relevance to enhance the likelihood of elaboration of the message (Petty and Cacioppo 1981a; 1986 and subsequent

I must confess that I am not satisfied with the subtlety of this evidence. Although the literature review I draw this information from was published last year (2010), the studies summarised were realised in the eighties, and thus could not anticipate recent theoretical evolutions in the field. A first layman remark is that the need for clear-cut experimental parameters leads to levelling complexity. In fact, many of these results reported are open to manifold interpretation. Behavioural psychologists alternate narration, pictures, music and semantic variants of the same term to induce emotion, but there is still room for research on the incidence of specific discourse 'genres' in emotion triggering (Blanc 2006). From the point of view of a rhetorician, the information that an upset person has poor results in mathematical tasks is of little use. Although D. Walton would probably find in this information a support for his view, namely that emotion is acceptable only when it is relevant to the objectives of the debate, in our opinion it is not easy to delineate relevant and irrelevant emotions, not least to do it objectively. Although distraction of attention by an emotion incidental to the task is an interesting phenomenon, mathematical calculations are very different from the texts which usually constitute a rhetorician's corpus of study. Besides, a rhetorician would tend to be more interested in the role of emotion as part of the experimental task, as such a role would echo much better the role of *pathos* in rhetoric.

My impression is that theories on emotion are much more advanced than models of observation. Researchers themselves agree that it takes time to incorporate theoretical insights into experiments and models (Blanc 2006). Finally, although experimental psychology does not explicitly adopt a normative stance on emotions, experiments echo views and prejudice of their designers over the concept of rationality. It will take some time for behavioural experimental psychology to incorporate the recent theoretical views of psychologists in models of observation. For the time being, it is difficult to assimilate the above mentioned findings in a rhetorical framework, because most of the times they are the product of a restricted vision of rationality, modelled on the computer metaphor of the mind. Although these studies consider emotion and logicity to be separate

publications). For a review of cognitive models of persuasion in advertising, see Derbaix & Grégory (2004).

variables, they do not confirm the view of the first wave of pragma-dialectics (1994) that emotions are fallacious (section 3 (i)). This is not surprising. These studies have an empiricist impetus and do not set themselves a model to confirm or a normative task to perform. Instead, they underline the importance of emotions to seek coherence and relevance and its involvement in high-order cognitive tasks.

### *4.3 Emotion and bias in the work of J. Elster*

A far more convincing interpretation of bias is proposed by the philosophy of mind. A cognitive-conative approach interprets these irrational emotions as an incongruent set of beliefs and desires (Elster 1983). We owe to this tradition the exploration of such phenomena as self-deception and wishful thinking. Such approaches are compatible with psychological theories that emphasise the activation of parallel mechanisms in emotion triggering. It is also possible to link their research topics to the interest of appraisal theories for goal-incongruent emotions.

I consider Elster's approach to be a balanced one, as it focuses on the role of emotions in indicating that values are at stake (2003) and not only in 'subverting rationality' in the traditional sense of it (1983). Elster maintains a terminology which opposes desires to beliefs, but allows for their cross-contamination. His account on emotion has evolved in time along the shift of paradigm presented in 4.4.

### *4.4. A paradigm shift in the view of Rationality*

A decisive shift in the study of emotion took place with Damasio's seminal work (1995) on emotionally impaired patients and with the publications on emotional intelligence that followed (Mayer and Salovey 1995, Goleman in Blanc 2006). The contemporary view of psychology on emotions is that they are necessary for social adaptation. Practical reasoning and value negotiation are part of it. Cognitive psychologists are now interested in the study of behaviour in relation with the attribution of subjective meaning to the environment (Blanc 2006: 66). New theories are in search for new models

and evaluative scales that do not associate emotion with logicity in the strict sense. This shift of paradigm probably accounts for the decline of experimental studies based on self-reporting. Rhetoric studies too advocate a paradigm shift that would lead to viewing rationality as encompassing emotion and subjectivity, in line with Damasio's work. For instance Dominicy (2001) underlines the propensity for action sought in epideictic praise. Today potential for action is consensually considered to be the evolutionary functionality of emotions.

In accordance with this shift of paradigm, some philosophers of mind account for the functionality of emotions in terms of their unique association with values. Some philosophers consider emotion as an equivalent of perception that takes some value as its *Intentional object* (Tappolet 2000, Livet 2002). According to Livet, emotions are indices of values we cherish; they can be shared socially and they motivate action. Resistance to modifying beliefs when new data imposes reconsideration is related to the association of these beliefs with values through emotions.

Having strong feelings about the way things ought to be or ought not to be (which is a core issue in argumentation) is far from being irrational, but may bring agents to irremediable disagreement. For instance, an appeal to pity may be irrelevant to the discussed topic for some participants in the debate, while other participants, though, may consider compassion to be a value that overrides other imperatives. Notably, sacrifice and self-abnegation cannot be accounted for by optimal choice or utilitarian models (Elster 2003, 2010). Disagreement over values is not an index of irrationality, or of sub-optimal judgment under the effect of these strong feelings, but a manifestation of incarnated preference.

Cognitive psychology and the philosophy of mind do not adopt a normative approach to emotion, one that would allow us to consider it *ex-ante* and by definition fallacious. They offer neither any straightforward criteria nor means to exclude mistakes and to differentiate between bias and subjectivity either. It seems, instead, that contemporary research focuses on subjectivity, on the functionality of mental operations (Blanc 2006: 63), and on their mutual shaping, rather than on emotion as a source of bias and



error.<sup>15</sup> What seems to evolve is the kind of tasks used to evaluate the links between emotion and cognition. This could be a symptom of a changing view of what should count as rational. The account of such a composite rationality which encompasses emotions confirms ancient rhetoric's intuition.

#### 4.5 *Swayed away, in spite of one's will?*

This short overview of the relation between emotion and cognition, either in terms of reciprocal influence or in terms of consubstantiality, according to the chosen paradigm does *not* entail any conclusion about the need (or uselessness) of ringing an alarm of manipulation because of the use of/appeal to emotion in discourse. Thus, we have given a negative answer to the first part of question (i) and we can deal with the second part of it: the impression of deceptiveness and emotions' impact.

Hence, I would like to turn to the triggering of emotion, in order to focus on two particular characteristics that seem to be widely accepted and may account for the problematic nature of emotion: automaticity and the possible absence of a propositional content.

Emotions are said to be triggered by beliefs, to have a cognitive antecedent that consists in the appraisal of a situation. Nevertheless, "appraisal is not knowledge" (Lazarus 1991: 152); it is considered to be a partially automatic process that "does not imply rationality, deliberateness or consciousness" (*ibid.*). This postulate reminds us of a well-known quandary of philosophy of mind about emotion, viz. the doubts on the existence of its Intentional object and on the status of its propositional content (Dominicy 2007). Philosophy of mind does not deal with cognitive mechanisms, but with the form of representations emotions consist in. It raises the question whether emotions that are endowed with Intentionality have a propositional

15 Talking about bias presupposes a possibility to choose between a biased and a non-biased version. To be able to talk about error (and about truth), we need a definition and a criterion (Dominicy 2005), even when knowledge of natural and institutional facts is concerned (Searle 1995). It does not make sense, though, to talk about error and truth in the case of epistemologically subjective experiences of objective realities (pain, joy) and in the case of preferences concerning deontic propositions ('freedom over equality').

content, like beliefs, or a non-propositional content, like perception. This dilemma is not so important for cognitive psychology, which conflates both types of Intentionality under the generic term of “representation” (Moors 2010: 27). But the philosophical distinction is important for the purpose of my intervention, as I believe that the putative absence of a propositional content and the partial automaticity of appraisal are the reason audiences may feel and be misled by skilful speakers, as I indicated under point (ii) above.

If compared to a fastidious and time-consuming evaluation of the same situation, appraisal is *imperfect*,<sup>16</sup> as it is closer, in reality, to a rapid overview and perception than to an exhaustive analysis and cannot be submitted to the standards of conscious judgement. Even if it were a conscious activity, appraisal would be a strenuous task. Appraising agents have to deal with lack of evidence, conflicting evidence, conflicting and mutually incompatible popular goals, and conflicting goal priorities, all these giving rise to appalling situations that can be responsible for uncertainty, risk and stress.<sup>17</sup> Again, as in the case of conflicting values, options are not definitely set, answers to questions are not clear-cut, and ambiguity reigns. Even if evidence is available and priorities straightforward, the means may lack and the alliances may be fragile. Contrary to what the computer metaphor suggests, the brain allows for ambivalence, non-conclusiveness and simultaneity of apparently incompatible beliefs and desires.<sup>18</sup>

To answer question (ii) we have to say that appraisal is not knowledge. As emotions involve automatic and parallel operations and, in some cases, are said to be devoid of propositional content, they may lead to what the agents themselves, with the benefit of hindsight, may consider as bias or

16 This is not to say that if agents took the time to precisely and explicitly evaluate goal congruence and their own capacity of control, the emotion could disappear.

17 Danblon (2002) associates persuasion with a feeling of safety, provided in the past by magicians and priests. In her conception of multilayered rationality, modern genres compensate for old needs through the ‘as if’ mechanism. Rhetorical convention is supposed to be perceptible for agents.

18 A subsidiary claim of my paper is that, at least in some cases, the search for straightforward stances and the cry against manipulation stem from a misrepresentation of the task to be performed. Not all problems have a solution, and not all tasks include problem-solving among cooperative peers. Conflict of interests and power struggles, ethical and political choices can fail to be resolved with the best argument.

unsatisfactory communication. Their fast and uncontrollable manifestation and their link with taking action expose agents to risks.

These observations bear consequences into the way an argumentation theory should provide for means of defence from what may be considered the subversion of an extended, composite rationality which encompasses both logicity and emotion.

If emotion is consubstantial to memory, attention and reasoning, if it is necessary for the vivid perception of incarnated values and social adaptation and if beliefs and desires cross-fertilise each other, there are not that many cognitive processes left for ‘cold knowledge’, certainly not practical reasoning which is mostly involved in argumentation. If emotion is widely present not to say inescapable, then we have to seek a way to come to terms with its presence in discourse instead of banishing it under the qualification of fallacious.

## 5. A pedagogical project for rhetoric

### 5.1. *Debate emotions for their own sake*

I would like, now, to turn back to rhetoric and argue in a quite provocative way for the idea that rhetoric in the broad sense of verbal situated communication to an audience or verbal interaction among interlocutors is not (only) the problem, the source of manipulative emotion, as it is sometimes portrayed to be, by rhetoric’s opponents, but (also) the remedy for it. I am echoing here C. Plantin’s (2009) heart-cry to “let people speak”.

Emotions can be valid reasons to act, if associated with social adaptation, survival and values. Furthermore, emotions can themselves be the centre of a debate, as their appropriateness is defined according to a culturally defined Topic – i.e., a set of rules and associations (Plantin 1998).<sup>19</sup> Appraisal

19 R. Micheli (2010) presents a corpus analysis of such a situation, where the legitimacy of emotions like pity and revenge was at the heart of the debate over the abolition of death penalty.

theorists develop such a topic about specific emotions, around ‘if-then’ sentences (Lazarus 1991). *Emotions are debatable*: speakers give reasons why they feel them, insert them in narrative scripts and challenge their legitimacy/suitability. But the fact that emotions are debatable does not imply that they are *effectively debated*.

We have seen that humans do have the cognitive capacity of empathy and self-knowledge, i.e., they can have representations of emotions – their own emotions and those of other people – but this does not mean that they systematically have such representations. To debate about emotions, one first has to be aware of them, and second to be able to put them in words. These steps are intertwined but not conceptually identical: interestingly enough, although the (im)possibility of thinking without language has been a philosophical conundrum, we can comfortably accept the idea of vague sensations and we are, culturally, less demanding in terms of precise expressions of subjective feelings. We saw that a great deal of emotion elusiveness is due to the automatic and possibly non-propositional character of the appraisal. Agents may perceive physiological reactions while a part of the appraisal representations are not explicit. In addition to that, an emotion can be felt without being recognised as such, specifically identified, labelled or expressed. The reasons why the process may remain incomplete vary: emotion awareness can be superfluous, time- and attention-consuming or painful, as not all emotions are socially acceptable and self-image flattering.

The good news is that humans are not at the mercy of their subjective understanding of reality. Communication promotes awareness and sets the conventional/social boundary for the acceptance of an emotion. When argumentation does not apply, fiction and poetry take over, to show what cannot be coped with otherwise. Communicators are sensitive to this conventional play with discourse ‘genres’.<sup>20</sup> Debating, questioning, acclaiming or refuting the adequacy of an emotion broadens the sphere of its propositional representation and promotes awareness and regulation of it. It is in this sense that discourse practice can offer a remedy to the problem of ‘manipulation through emotion’ that it is supposed to create.

20 Danblon (2006) stresses the importance of the conventional ‘as if’ mechanism, which allows for fiction to play the comforting role once held by magic.

The need to develop defensive skills and attitudes in communication is a reason to train young people in rhetoric, to raise awareness over potential pitfalls and our inescapable exposure to them due to the way our brain works. Familiarisation with rhetoric's tools and practice is the best way to prevent some unsatisfactory outcomes in communication. This familiarisation does not mean proscribing analytical categories or demonizing emotions a priori, or constructing a theory of manipulation, but training people to defend themselves in communication itself. An example of the way I perceive such a defence is illustrated in section 5.2. As a matter of fact this is the way people do communicate: all I am suggesting is that the more they practice, the better they cope with it.

As a rapid appraisal of situations, emotions may lead us to conclusions we may revise and acts we may regret with the benefit of hindsight. The risk diminishes as our knowledge of our self and of other people rises. This risk is the price to pay for a mechanism which brings us in contact with our human nature, our positive and negative values, which are a crystallisation of past collective experience.

### *5.2. Does a descriptive vocation equal permissiveness?*

Now, how about permissiveness and standards?

Although rhetoricians typically abstain from reasoning in terms of manipulation and fallacies, in practice they use analytical categories that may inspire a critic. I would like to associate two of these categories to the use of emotion in speech. The first one is the Aristotelian triangle of *logos*, *ethos*, *pathos*. The second one is the so called 'genre'.

J.M. Adam considers manipulation to be the relative hypertrophy of *ethos* or *pathos* to the detriment of *logos* (2002: 46). Although he does not propose this view as a formal definition of manipulation, a term he rather uses once in its non-technical occurrence, I would like to extend his view and use it myself as a functional definition/criterion to talk about *potential* manipulation (in the non-technical occurrence in question (ii)) in the case of the

hypertrophy of one of the three proofs to the detriment of the others).<sup>21</sup> ‘Genre’ can also help us to come to terms with emotion in speech. We can expect a higher amount and variety of emotion to be present in fiction than in an essay on a new molecule. Still, there is much room for hybrid genres and the introduction of subjectivity, even in science.

These functional criteria are indicative rather than conclusive about the existence or not of ‘manipulation through emotion’ in a specific speech. Nevertheless, the absence of a systematic conclusive theory on ‘manipulation through emotion’ in rhetoric does not mean ‘permissiveness’. It is rather an indication of a view of the role of the expert and of the way a debate is closed.

The researcher in rhetoric does not usually take a normative stance towards the speech as he considers that the outcome of a debate has to be decided by the participants and not to be proclaimed by a linguist-expert. It is a matter for the participants as well to decide on the appropriateness of the use of emotion, which is a metalinguistic topic among others to discuss in a specific debate. To give an example, statements (a) to (c) of section 3 could respectively be dismissed in the following way:

- (a) So tempting! Can I afford it? / Do I really care for status? / The salesman is trying to do his job, but he won’t get me.
- (b) In a world of free movement of capitals, we have to allow for some movement of people too. / You are inciting hatred. / Who else wants to become a waiter nowadays?
- (c) We’ll beat you! / Nationalism in our region has caused two world wars and a regional one. Stop warmongering!

The above examples are artificial ones for the purposes of this article. In natural debates a refutation may take more than an exchange of a couple of sentences. Argumentative chains in verbal interactions are open to further refutations, not least when the topic is interwoven with clashes of power and interest and/or ethical options. The fact that sometimes the need to take a decision seems to close a matter (ideally by vote) does not mean that

21 A counter-intuitive illustration of this idea would be a technocratic report about grants to the banking sector, set in an impersonal, emotionless mode of supposedly pure *logos*. Rhetoricians will recognise a variation of *ad verecundiam* in this evanescence of *ethos* in the name of impartiality. An implicit appeal to the expert opinion through an impersonal style gives an impression of objectivity.

theoretically it cannot be reopened if new elements appear and somebody assumes the burden of proof. In such a model of ‘*self-regulation*’ and *through an endless possibility of refutation*, the rhetorician may opt for participating in a debate, without the claim to close it, as if he were a judge or a referee.

There are also cases where a rhetorician may feel strong on the stakes in a debate and where, as a result, he will unavoidably take stance. The correctness and efficiency of this stance though should also be a matter for the participants in the debate to tell, as they do not automatically stem from a better knowledge of the argumentative tools a linguist may have.

Although the educational project rhetoric could have should help the participants in a debate to be more satisfied with the outcome of this debate, it has in itself its own limits. Ethical dilemmas, conflict and subjectivity (expressed through emotion) are inherent in numerous debates. These debates may be opened, closed or reopened, but unlike problems, they cannot be ‘solved’.

## 6. Conclusion

Some theories of argumentation consider appeals to emotion to impair critical reasoning. On the opposite side, rhetoric studies *pathos* as a natural and legitimate means of proof. I tried to explore if the same divide can be found in cognitive science. Although in the past emotions and logicity have been studied as two separate variables, contemporary theories in cognitive psychology perceive emotion and cognition as mutually shaping each other, indicating that ‘cold knowledge’ may be an illusion. Furthermore they consider emotion to be vital for survival, social adaptation, perception of values and coping with tasks requiring subjective evaluation. This extended view of rationality confirms ancient rhetoric’s intuitions.

Nevertheless, as emotions involve automatic and parallel operations, a rapid appraisal of the situation and, in some cases, are said to be devoid of propositional content, they may lead to what the agents themselves with the benefit of hindsight may consider to be bias or unsatisfactory outcome of the communication. These observations bear consequences into the way

argumentation theory should perceive rationality and provide for means of defence from what may be considered the subversion of it. In an extended view of rationality, which includes emotions, these are not necessarily contrasted to logicity, but to other emotions that are more suitable to the situation according to the participants in a debate.

Since emotions are embedded in our rationality (Damasio 1995, Blanc 2006), considering them fallacious is a vain struggle on the basis of a misperception of the mind. I propose to deal with emotions through 'self-regulation' in discourse practice as an alternative option to the *ex-ante* definition of acceptance criteria and conditions in a system. Emotions can be shown, told, debated and shared *ex post* through numerous discursive 'genres' and modes. Albeit debatable, emotions are not always effectively debated. Communication broadens the sphere of propositional representations of emotions by offering a natural remedy against 'manipulation through appeal to emotion'.

Rhetoric does not adopt a normative stance towards manipulation in accordance with its particular view of the linguist's role and of the way a debate is closed. Nevertheless it can contribute to diminish the risks of manipulation through a pedagogical project of familiarisation of audiences with rhetorical tools. The better we know ourselves, our positive and negative values and the tools speakers use to convince, the less we are exposed to the risks of 'manipulation'.

Furthermore, the claim against 'manipulation through appeal to emotion' is a legitimate metalinguistic argument in a debate, although its acceptance depends on the participants in the same debate. In the self-regulatory model described above, the participants and not an external expert/linguist get to decide whether they deem some emotions to be relevant and appropriate in a specific debate and whether they want to accept such a metalinguistic argument.



## References

- Adam, J.-M. (2002): “De la grammaticalisation de la rhétorique à la rhétorisation de la linguistique: aide-mémoire”, in Koren, R. & Amosy, R. (eds), *Après Perelman: quelles politiques pour les nouvelles rhétoriques?* L’Harmattan, Paris, 23-55.
- Adam, J.-M. & Heidmann, U. (eds) (2005): *Sciences du texte et analyses du discours. Enjeux d’une interdisciplinarité*. Slatkine Érudition, Genève.
- Anscombe, J.-C. & Ducrot, O. (1983): *L’argumentation dans la langue*. Mardaga, Bruxelles.
- Aristote (1991): *Rhétorique*. Le Livre de poche, Paris.
- Blanc, N. (2006): *Émotion et cognition*. Éditions In Press, Paris.
- Blanchette, I. & Richards, A. (2010): “The influence of affect on higher level cognition”, in De Houwer, J. & Hermans, D. (eds), *Emotion and Cognition: Reviews of current research and theories*. Psychology Press, Hove, UK, 561-595.
- Berne, E. (1996) [1964]: *Games people play*. Ballantine Books, New York.
- Bouvier, A. (1995): “Les paralogismes d’un point de vue sociologique”, *Hermès* 16, 45-55.
- Clément, F. (2006): *Les mécanismes de la crédulité*. Droz, Genève / Paris.
- Damasio, A.R. (1995): *L’erreur de Descartes. La raison des émotions*. Trad. par. M. Blanc. Odile Jacob, Paris.
- Danblon, E. (2002): *Rhétorique et rationalité, Essai sur l’émergence de la critique et de la persuasion*. Éditions de l’Université de Bruxelles, Bruxelles.
- (2005): “Discours magique, discours rhétorique”, in Adam, J.-M. & Heidmann, U. (eds), *Sciences du texte et analyses du discours. Enjeux d’une interdisciplinarité*. Slatkine Érudition, Genève, 145-160.
- De Houwer, J. & Hermans, D. (eds) (2010): *Emotion and Cognition: Reviews of current research and theories*. Psychology Press, Hove, UK.
- Derbaix, C. & Grégory, P. (2004): *Persuasion. La théorie de l’irrationalité restreinte*. Economica, Paris.
- Dominicy, M. (2001): “L’épidictique et la théorie de la décision”, in Dominicy, M. & Frédéric, F. (eds), *La mise en scène des valeurs. La rhétorique de l’éloge et du blâme*. Delachaux et Niestlé, Lausanne / Paris, 49-78.
- (2005): “Langage, interprétation, théorie. Fondements d’une épistémologie moniste et faillibiliste”, in Adam, J.-M. & Heidmann, U. (eds), *Sciences du texte et analyses du discours. Enjeux d’une interdisciplinarité*. Slatkine Érudition, Genève.
- (2007): “Sémantique et philosophie de l’esprit: les rapports de perception visuelle”, in Neveu, F. & Pétilion, S. (eds), *Sciences du langage et sciences de l’homme. Actes du colloque 2005 de l’Association des Sciences du langage*. Lambert-Lucas, Limoges, 65-82.
- Dominicy, M. & Frédéric, F. (eds) (2001): *La mise en scène des valeurs. La rhétorique de l’éloge et du blâme*. Delachaux et Niestlé, Lausanne / Paris.
- Eemeren, F.H. van & Grootendorst, R. (1996): *La nouvelle dialectique*. Kimé, Paris.
- Eemeren, F.H. van & Houtlosser, P. (2006): “Strategic Manoeuvring: a synthetic recapitulation”, *Argumentation* 20, 381-392.
- Elster, J. (1978): *Ulysses and the Sirens*. Cambridge University Press, Cambridge.

- (1983): *Sour grapes: Studies in the subversion of rationality*. Cambridge University Press, Cambridge.
- (2003): *Proverbes, maximes, émotions*. PUF, Paris.
- (2010): *Le désintéressement. Traité critique du monde économique*. Editions du Seuil, Paris.
- Harris, M. (1976): “History and significance of the EMIC/ETIC distinction”, *Annual Review of Anthropology* 5, 329-350.
- Koren, R. & Amossy, R. (eds) (2002): *Après Perelman: quelles politiques pour les nouvelles rhétoriques?* L’Harmattan, Paris.
- Lazarus, R.S. (1991): *Emotion and Adaptation*. Oxford University Press, New York.
- Livet, P. (2002): *Émotions et rationalité morale*. PUF, Paris.
- Micheli, R. (2010): *L’émotion argumentée*. Éditions du Cerf, Paris.
- Moors, A. (2010): “Theories on emotion causation: a review”, in De Houwer, J. & Hermans, D. (eds), *Emotion and Cognition: Reviews of current research and theories*. Psychology Press, Hove, UK.
- Pareto, V. (1968) [1916]: *Traité de sociologie générale*. Droz, Genève.
- Plantin, C. (1995): “L’argument du paralogisme”, *Hermès* 15, 245-262.
- (1998): “Les raisons des émotions”, in Bondi, M. (ed.), *Forms of Argumentative Discourse*. CLUEB, Bologne, 3-50.
- (2002): “Analyse et critique du discours argumentatif”, in Koren & Amossy, *Après Perelman: quelles politiques pour les nouvelles rhétoriques?* L’Harmattan, Paris, 229-263.
- (2009): “Laisser dire: évaluer c’est encore argumenter. De la surabondance de règles sur la parole”, in *Séminaire doctoral. Modèles de l’argumentation*, ULB, 9 décembre 2009.
- Plantin, C., Doury, M. & Traverso, V. (eds.) (2000): *Les émotions dans les interactions*. Presses Universitaires de Lyon, Lyon.
- Searle, J.R. (1995): *The Construction of Social Reality*. Penguin Books, Londres.
- Tappolet, C. (2000): *Émotions et valeurs*. PUF, Paris.
- Toulmin, S.E. (2003) [1958]: *The Uses of Argument*. Cambridge University Press, Cambridge.
- Walton, D. (1992): *The place of emotion in argument*. The Pennsylvania State University Press, University Park, Pennsylvania.
- Woods, J. (2004): *Argument, critical thinking, logic and the fallacies*. Prentice-Hall, Toronto.

# L'argument d'autorité: de sa structure à ses effets

Thierry HERMAN, Universités de Neuchâtel et de Lausanne

## 1. Introduction

«Le principe de l'argument d'autorité est si bien analysé qu'il semble difficile d'apporter à sa description quelque élément nouveau» (Doury 1999: 1, online). L'argumentation cherchant une légitimation par un propos émanant d'une autorité est un schème fréquent, son fonctionnement semble assez limpide a priori et son efficacité est souvent démontrée (cf. par exemple Cialdini 2006: 208 sqq). En outre, il est à lui seul une des quatre familles d'arguments dans la typologie proposée par Philippe Breton (2005) et figure en bonne place – sinon en première place – dans les manuels d'introductions aux sophismes de logique informelle (Tindale 2007, Walton 2006 par exemple). Dans la tradition normative de la logique informelle, le schème argumentatif se repère par la co-présence explicite d'une proposition énoncée P et de son locuteur X. Selon Tindale (2007), parmi d'autres manuels d'argumentation de la même ligne de pensée, on peut en effet résumer l'argument d'autorité de la manière suivante: «Quelqu'un (ou quelque source) affirme un propos P. Donc P est vrai» (2007: 131, notre traduction).

Implicitement, l'argumentation implique une prémisse selon laquelle la source en question est une autorité au sujet de P.

Nous voudrions montrer que l'argument par l'autorité demande d'être précisé par rapport à une définition commune, présente une structure qui n'a rien de commun avec la très grande majorité des autres schèmes argumentatifs et implique l'idée d'imposer une réalité qui peut être problématique à appréhender lorsque cette réalité est imposée par l'autorité du locuteur. Nous voudrions plus spécifiquement illustrer dans cette contribution les points suivants:

1. Il nous semble que la conclusion de l'argument d'autorité touche moins à une question de vérité de l'énoncé – ce qui serait typique d'une tradition logique – qu'à une question de certitude (assumée par le

- locuteur [cf. section 2.2. et 3.1]) et d'indubitabilité postulée – notions observées ici dans une dynamique dialectique (cf. section 3.2.).
2. A ce titre, la structure même de l'argument d'autorité, atypique, paraît servir une tentative d'imposer P à l'allocataire comme étant indubitable (cf. section 2)
  3. L'argument d'autorité a souvent été considéré comme un sophisme dans le cas du recours à une fausse autorité ou à une autorité non qualifiée. Comme d'autres, nous pensons impossible de le considérer d'emblée comme tel. Et nous nous demanderons même jusqu'à quel point l'analyse critique du schème est possible (cf. section 1).
  4. Selon la structure proposée par Tindale, la source X est explicite et l'idée qu'elle soit experte dans le domaine peut être – et est souvent – implicite. Cela incite à une triple réflexion. (i) Comment traiter un cas de discours rapporté du type «Selon Schmuingh, il fera beau demain» si «Schmuingh» est parfaitement inconnu de l'allocataire? (ii) Comment traiter le cas où la source est posée comme existante par un indice d'évidentialité, mais non identifiable: «Il paraît qu'il fera beau demain?» (iii) Comment traiter le cas où la source n'est même pas mentionnée: «Il fera beau demain»? On peut penser que ce dernier cas ne touche plus à l'argument d'autorité; pourtant, il partage une forme de certitude épistémique et une tentative d'imposer P comme indubitable. Ne pourrait-on pas considérer ceci comme une forme de rhétorique de l'autorité dont l'argument d'autorité représenterait une forme particulière? (cf. sections 3.3 et 3.4)

## 2. Un sophisme inévitable

De manière générale, plusieurs sophismes sont aujourd'hui remis en question en tant qu'ils sont caractérisés comme tels. Didier Maillat évoque dans ce volume des «sophismes qui ne sont pas si fallacieux». La remise en question de ce qui est juste ou faux, acceptable ou inacceptable peut se lire sous la plume de Marc Angenot: «C'est ce que je montrerai de *tous* les prétendus sophismes: tous sont matière à débat et aucun ne se disqualifie nettement et

indiscutablement» (2008: 167). Une autre contestation, plus étonnante au regard de l'abondante littérature que cet auteur a produite sur les sophismes, émane de Douglas Walton (avec Fabrizio Macagno et Chris Reed), lequel, dans *Argumentation Schemes* (2008), doute que les sophismes doivent encore être traités comme tels et propose de remplacer le terme de *fallacies* par des schèmes argumentatifs défaisables au moyen de questions critiques.

Un des sophismes qui a fait vaciller le plus l'approche normative a priori largement pratiquée dans les courants de logique informelle ou de pensée critique (Herman 2011) est sans doute l'appel à l'autorité. Celui-ci est encore fréquemment considéré comme un sophisme:

L'appel à l'autorité a traditionnellement été traité comme un sophisme par les manuels de logique. [...] Souvent, ils ne distinguent pas clairement différentes sources d'autorité et, peut-être influencés par la vue de Big Brother selon laquelle l'autorité centrale dogmatique est le grand problème, ont tendance à considérer l'appel à l'autorité comme généralement fallacieux. (Walton 1997: 28, notre traduction)

Or, comme le dit Tindale dans son manuel: «lorsque je consulte un médecin ou un mécanicien, je ne pense pas que les délibérations qui en découlent soient fallacieuses» (2007: 128, notre trad.). De même, Walton affirmait déjà en 1997 que «la plupart des choses que nous acceptons sont, inévitablement, acceptées sur la base de l'autorité» (1997: 1, notre trad.). Jean Goodwin affirme tout autant que les appels à l'autorité sont «inévitables» (1998: 267), rejoignant ainsi Charles Willard pour qui toute prise de décision publique est «dépendante de l'avis d'experts faisant autorité» (1990).

En ce sens, l'appel à l'autorité est un mouvement argumentatif qui tend à imposer une proposition P, la plupart du temps sans que l'allocutaire ait les moyens ou les compétences de remettre P en question. Cette nécessité de croire sans vraiment savoir peut susciter quelque gêne: idéalement, nous aimerions nous passer de l'appel à l'autorité, car, selon Walton, «nous aimons penser que nous pouvons nous forger des croyances par nous-mêmes et de manière autonome, en nous réservant le droit d'être sceptiques [...]» (Walton 1997: 1). Un tel idéal n'est bien sûr pas réaliste. Comme Wittgenstein, nous devons la plupart du temps nous résoudre à admettre que: «Ma vie revient à me contenter d'accepter beaucoup de choses» (*On Certainty*, cité par Angenot 2008: 203-204). En ce sens, l'appel à l'autorité, par son caractère inévitable que les autres sophismes ne partagent pas, occupe une place à part parmi les arguments potentiellement fallacieux.

Le mariage entre l'idée de sophisme et un raisonnement incontournable ne paraît pas des plus harmonieux. Selon la tradition, le respect dû aux autorités brouillerait les capacités de raisonnement et conduirait à accepter des vérités frelatées au nom de ces autorités. Le problème est ontologique: le sophisme constitue une forme d'écart, une sortie de route, par rapport à la ligne droite de la raison. Le recours aux sophismes méritant les opprobres dans certaines écoles de pensée, il ne faudrait idéalement pas céder à la tentation de l'*ad verecundiam* et expurger tout discours de cet appel à l'autorité.<sup>1</sup> Or, nonobstant le caractère captieux de ce type d'argumentation, l'appel à l'autorité, si Walton et d'autres disent vrai, forge la plupart de nos connaissances, de nos certitudes, de nos opinions. Ce qui se produit empiriquement est donc tout le contraire. La connaissance acquise sans recours à l'autorité constituerait de fait l'exception plutôt que la règle. Comme le dit Haskell, «nous croyons à l'évolution, non parce que nous avons en tête les preuves et les expériences qu'il faudrait faire pour comprendre tout le processus et le saisir de manière rationnelle, mais parce que nous nous fions aux biologistes» (Haskell 1984: xi, cité par Willard 1990: 1, notre trad.). De sorte que l'on peut légitimement se demander pourquoi l'appel à l'autorité, qui est non seulement fréquent, mais surtout nécessaire, figurerait dans la longue liste des *fallacies* dressée par la logique informelle. Depuis peu, plusieurs théoriciens admettent l'*ad verecundiam* comme potentiellement légitime. Selon Walton, les manuels des années 60 et 70 faisaient tous de l'appel à l'autorité un sophisme, avant d'admettre plus récemment que cet appel peut être raisonnable en certaines circonstances (1997: 28-29).

On peut prendre comme signe de cette évolution de la pensée l'exemple du site anglais *fallacy files*, qui catalogue dans sa liste de sophismes l'appel à une autorité trompeuse (*misleading authority*). À défaut de ne pouvoir écarter le recours aux experts comme une technique argumentative somme toute acceptable et efficace, on propose de mettre en question la qualité des experts convoqués à l'appui du raisonnement en ajoutant une épithète au schème argumentatif. Cette désignation d'appel à l'autorité trompeuse représente plus qu'une simple substitution d'un nom par un autre. Avec cette mention, l'*ad verecundiam* n'est plus lié à l'argumentation, au sens de la matérialité

1 L'argument *ad verecundiam* – littéralement un appel au respect dû à l'autorité – est une désignation forgée par John Locke qui constitue à nos yeux une forme particulièrement restreinte de l'appel à l'autorité, celui de l'autorité de prestige (voir plus loin).

textuelle que cela représente, mais à la situation rhétorique à chaque fois particulière dans laquelle s'inscrit l'appel à l'autorité. Ce qui est en jeu est moins une question de degré de confiance qu'il faudrait accorder à l'argument proprement dit qu'une question de crédibilité de l'autorité citée. En soi, l'appel à l'autorité ne serait plus blâmable en tant que mouvement argumentatif pour autant que l'autorité en question apporte des gages de rectitude morale, scientifique, etc.

Ce faisant, il faut bien mesurer que l'on remplace une impossibilité par une difficulté qui n'est pas moindre. À l'impossibilité matérielle de conduire les expériences scientifiques plutôt que de se fier aux experts (sans parler de l'autorité du témoignage qui est non reproductible), on substitue l'extrême difficulté de jauger avec certitude la crédibilité de l'expert cité. Pour la plupart des autres sophismes, il est possible de démontrer l'erreur logique – par exemple l'affirmation du conséquent –, il est possible de dénoncer un appel trop massif aux passions ou de montrer le caractère probablement trop généralisant d'une induction en prenant principalement appui sur la matérialité textuelle. Mais l'autorité trompeuse a ceci en commun avec le mensonge par omission qu'elle nécessite une certaine connaissance de la situation rhétorique (au sens de Bitzer 1968) pour être découverte et contrée. La matérialité textuelle n'offre guère d'appui. On trouvera difficilement des listes d'experts *grata* et *non grata* et, quand elles existent, elles sont autant contestées que contestables en fonction du lieu social où l'on se trouve: une figure scientifique peut être révéérée par certains et honnie par d'autres. À défaut de liste, on sait bien que l'on utilise des heuristiques plus ou moins fiables, en se fiant à des indices de crédibilité qui ont été largement testés par ailleurs (cf. dans ce volume, la notion de vigilance épistémique, ainsi que plusieurs autres contributions [Macagno, Oswald et Lewinski, de Saussure] interrogeant les conditions qui poussent un allocutaire à admettre une proposition P comme étant vraie).

Les manuels d'argumentation mettent en scène différents cas de figure et demandent d'appliquer quelques principes de précaution. Par exemple, lorsque le footballeur et acteur Eric Cantona parle économie, la ministre française de l'Économie Christine Lagarde peut facilement récuser l'autorité: «Je crois que quelqu'un qui est un grand footballeur ou un grand acteur de cinéma doit se garder d'intervenir dans le domaine financier, économique,

surtout quand il n'en maîtrise pas les mécanismes» (1<sup>er</sup> décembre 2010).<sup>2</sup> Mais ce processus d'illégitimation n'a rien d'absolu et peut figer les acteurs de la vie sociale ou politique dans des rôles qui ne sont peut-être pas ou plus les leurs. Et si Eric Cantona avait dans sa retraite de footballeur étudié de près la macro-économie mondiale? Autre exemple: si un gourou de secte affirme A, plusieurs manuels d'argumentation proposeraient vraisemblablement qu'A soit a priori rejeté sans autre forme d'examen comme éminemment suspect, non crédible, inepte, peut-être indépendamment du fait que même un gourou de secte peut parfois avoir des propos sensés. Walton va jusqu'à proposer un examen critique de l'autorité qui introduit la question des domaines d'expertise. Il ne suffit pas en effet à ses yeux d'être légitimé par le corps social comme un expert, encore faut-il l'être dans le domaine qui concerne l'affirmation A. Citons, pour être plus clair, les six questions critiques pour défaire un appel à l'autorité:

1. Le jugement défendu par l'autorité tombe-t-il dans le champ de compétences dans lequel cet individu est un expert?
2. L'autorité citée l'est-elle pour son statut d'expert et non pour son prestige, sa popularité ou son statut de star?
3. Quel est le degré d'autorité de l'expert particulier? Même si l'individu cité est un expert légitime dans le champ dans lequel se situe la question traitée, la force avec laquelle l'appel doit être pris comme un argument plausible doit être mise en question.
4. Y a-t-il des désaccords parmi les autorités qualifiées consultées?
5. Les éléments de preuve objectifs sont-ils disponibles et si oui, en accord avec l'expert?
6. Les propos de l'expert ont-ils été bien compris? (d'après Walton 1997: 211, traduction libre)

On le voit bien, les conditions draconiennes livrées ici pour que le recours à l'autorité échappe à la désignation de sophisme paraissent pour le moins irréalistes *in situ*. En effet, répondre à ces six questions requerrait de toute évidence un degré d'expertise de l'allocutaire qui serait pratiquement équivalent à celui de l'autorité citée: comment consulter plusieurs experts,

2 <http://www.lefigaro.fr/conjoncture/2010/12/01/04016-20101201ARTFIG00538-la-revolution-cantona-irrite-christine-lagarde.php>



être sûr d'avoir compris l'expert cité, savoir son domaine de spécialité, connaître sa réputation dans ce domaine, sans être au moins bien informé, sinon soi-même un spécialiste du domaine (voir aussi Kutrovatz 2011)? Dans pareil cas, le rôle persuasif de l'autorité convoquée se trouverait réduit à une forme de biais de confirmation, l'expert apportant un point de vue que l'allocutaire partage déjà.

Si la question de la crédibilité de l'autorité est effectivement essentielle dans le processus de persuasion, cette crédibilité est, d'après les psychologues de la persuasion, moins jugée rationnellement par la série des six questions critiques de Walton que jaugée par des règles non écrites que les anglophones appellent *rules of thumb*, règles rapidement appliquées et non sans faille.<sup>3</sup> Pour reprendre l'exemple rapidement mentionné de l'écrit scientifique, Willard mentionne avec un malin plaisir dans cette pratique discursive la multiplication des citations d'auteurs entre parenthèses, ce qui permet d'évacuer la question de la définition des termes employés – on renvoie au travail d'autrui –, mais comporte aussi l'avantage de valoir pour preuve sans autre forme d'examen prévue. Plus le nombre d'auteurs est cité par un processus bien connu de *name-dropping*, moins le lecteur a des chances d'enquêter sur ces citations sauf naissance d'un puissant soupçon, ajoute encore Willard (1990: 12). Le coûteux travail de vérification des faits, voire de vérification du crédit scientifique de noms propres mentionnés, ne se fait que dans des conditions bien particulières et se révèle donc en pratique presque aussi difficile à réaliser que la vérification des données livrées par une autorité.

En fin de compte, même si on déplace le sophisme de l'appel à l'autorité vers l'appel à l'autorité fallacieuse, il reste difficile de se prémunir contre ce type de mouvement sans prendre des décisions liées aux contingences particulières de chaque situation. Cet état de désarmement devant le recours à l'autorité est d'autant plus amplifié que la forme même de l'argument, comme on le verra dans notre section suivante, ne facilite pas l'accès à un point de vue critique à l'encontre de l'autorité.

3 Les programmes de recherche sur les heuristiques cognitives travaillent intensément ces phénomènes. Oswald et Hart suggèrent que les questions critiques de Walton, ou ce qui y ressemble intuitivement, sont intégrées dans nos filtres de vigilance épistémique (Oswald & Hart à paraître).

### 3. Une structure atypique et ses effets stratégiques

Dans l'approche «traditionnelle» de l'argumentation, ce qui définit une argumentation est une structure ou une «composition d'énoncés» comme le dit Plantin (à paraître). Spécifiquement, «une argumentation est un discours où un énoncé, l'argument, appuie un autre énoncé, la conclusion» (Plantin, à paraître). C'est par exemple dans ce sens que les courants anglo-saxons de la logique formelle et informelle traitent l'argumentation. Tant les arguments que les conclusions peuvent être inférés à partir du contexte sans être explicitement présents, mais il y a toujours l'idée d'une construction d'un point de vue à partir d'un autre.

Or, l'argument d'autorité ne s'inscrit pas au premier plan dans une vision classique ou structurelle de l'argumentation. Ce raisonnement agit plutôt au second plan d'une structure plus complexe.

#### 3.1 *Structure ramassée de l'argument d'autorité*

Rappelons d'abord la définition de ce schème argumentatif. Dans le schème de l'argument par une source experte, tel que le proposent Walton, Reed et Macagno (2008: 310), l'argumentation se présente sous la forme suivante: «La prémisse majeure: la source E est experte dans le domaine S qui contient la proposition P; la prémisse mineure: E affirme que la proposition P est vraie ou fausse; la conclusion: P est vrai (ou faux).»

Aussi rigoureuse que soit cette définition, elle pose un problème empirique. Il faut en effet bien constater que la conclusion de la forme syllogistique proposée ne se rencontre pratiquement pas dans des discours authentiques, hormis dans certains exemples ironiques comme «si Aristote le dit, c'est que ça doit être vrai». Il ne s'agit pas seulement du problème classique de l'enthymème pour lequel il faut reconstituer des prémisses implicites, ce qui est le cas de la majeure partie des argumentations. En effet, l'enthymème, de manière générale et préférentielle, laisse souvent implicite une prémisse plutôt qu'une conclusion alors que dans le cas de l'argument d'autorité, on peut généralement observer empiriquement deux opérations simultanées:

a. *La concaténation des deux prémisses*: dans l'énoncé, «Selon les prévisionnistes, le beau devrait faire son retour demain», la prémisse «les prévisionnistes sont des experts en météorologie» est implicitement déclenchée par la mention de la source et nos connaissances empiriques sur le métier de prévisionniste. A titre de comparaison, on peut se demander si «Selon Paul, il pleuvra demain» convoque la prémisse de l'expertise de Paul. Hors cotexte plus large, il est sans doute difficile de le postuler. Mais nous y reviendrons.

Indépendamment de cela, l'observation de cas atypiques, où la proposition P et l'autorité sont présentées en deux énoncés, permet de mettre en relief le caractère inattendu de P. Voici un corpus non-représentatif d'exemples où nous avons cherché online et dans une base de données de presse écrite ([www.smd.ch](http://www.smd.ch)) des occurrences de l'expression «(et) c'est un expert qui»:

- (1) Le réchauffement climatique est en pause depuis dix-sept ans. Et c'est un expert qui nous le dit, le président du Groupe intergouvernemental sur l'évolution du climat (Giec), Rajendra Kumar Pachauri.  
(<http://www.scopalto.com/causette/35/halte-a-l-enfumage>)
- (2) «On peut faire avaler n'importe quoi aux gens.» Et c'est un expert qui le dit: Marcel Duchamp.  
(<http://blablartcontemporien.wordpress.com/category/le-cas-marcel-duchamp>)
- (3) «La mayonnaise a pris, c'est sûr.» C'est un expert qui le dit. Charly Pleux-Decollogne fait figure d'ancien à Lausanne puisqu'il entame sa troisième saison à la Vallée. (24 heures, 03.10.2008)
- (4) Ne pas mettre de limites à un enfant: voilà la vraie maltraitance! C'est un expert qui le dit. Ancien directeur adjoint du Service cantonal valaisan de la jeunesse, le psychologue Maurice Nanchen le répète. (L'Express / L'Impartial, 03.10.2008)
- (5) La grogne des Genevois à propos du nouveau réseau des TPG est totalement justifiée. C'est un expert qui le dit: professeur de sociologie urbaine et de mobilité à l'École polytechnique fédérale de Lausanne (EPFL), Vincent Kaufmann affirme que le système mis en place sert beaucoup plus les intérêts des TPG que ceux des usagers. (Tribune de Genève, 30.12.2011)

Les exemples 4 et 5 ont ceci d'intéressant qu'ils figurent en tête de l'article journalistique. La dimension provocatrice des énoncés avancés en guise d'accroche du lectorat est ainsi soulignée avant d'être tempérée par son caractère «autorisé». Mais les autres exemples livrent pareillement un énoncé

potentiellement contestable, sinon polémique, que le scripteur, dans un second temps, va rendre légitime par une autorité. L'absence de concaténation entre la prémisse fixant l'autorité et la proposition défendue pourrait s'expliquer par le caractère para-doxal de la proposition elle-même et la volonté de précisément provoquer;<sup>4</sup> on perdrait cette dimension dans la reformulation par exemple de (4) en: «Le psychologue Maurice Nanchen affirme que ne pas mettre de limites à l'enfant est une vraie maltraitance» - sans perdre pour autant le mécanisme même de l'argument d'autorité. Nous soulignerons toutefois la différence dans l'effet: la concaténation de prémisses pose une proposition comme peu contestable alors que sa séparation – lorsque la source experte est précisée dans un énoncé ultérieur à la proposition P – joue sur la contestation possible ou attendue de P avant que celle-ci ne soit en quelque sorte jugulée par la mention de l'autorité.

b. *L'absence de la conclusion*: on perçoit bien qu'il serait absurde pour un présentateur de météo de faire suivre l'énoncé P «Selon les prévisionnistes, le beau devrait faire son retour demain» par Q «Donc il fera beau demain». La vérité de Q est redondante par rapport à P. Dans l'exigence de progression de l'information étudiée en linguistique textuelle, Q n'apporte guère d'information nouvelle par rapport à P. La conclusion du syllogisme de l'appel à l'autorité joue en fait plutôt un rôle de confirmation d'une vérité au fond déjà assumée implicitement. En effet, des travaux tant en philosophie qu'en psychologie illustrent la clairvoyance du philosophe Thomas Reid (2000 [1764]) lorsqu'il exposait son «principe de véracité» et son «principe de crédulité» – le premier est «la propension à dire la vérité et d'utiliser le langage pour exprimer nos vrais sentiments» alors que le second «renvoie à la propension à se fier à la véracité des autres et «à croire ce que les autres nous disent» (Lehrer 1989: 75; Clément 2006: 61). Autant Ruth Millikan (1987) que Tyler Burge (1993) ont abondé en ce sens; Burge parle d'un principe

4 L'exemple (5) paraît plus complexe, dans la mesure où la grogne des Genevois paraît générale et donc plus refléter une doxa que la provoquer. La provocation vient peut-être du ton démagogique adopté ici et/ou de la force d'une opinion qui doit plutôt illustrer un genre du commentaire qu'un genre de l'information. Notons que l'absence de guillemets prolonge cette ambiguïté: si le locuteur de cette première phrase est un journaliste, comme on pourrait le penser, la provocation est dans la rupture par rapport aux normes d'écriture journalistique. Tout se passe comme si l'irritation du journaliste lui faisait oublier son devoir de neutralité... Ce que la deuxième phrase va bien entendu désamorcer.

d'acceptation: «Une personne est incitée à accepter comme vrai quelque chose qui est présenté comme vrai et qui lui est intelligible, à moins que de fortes raisons l'en empêchent» (1993: 467). En psychologie sociale, les travaux de Gilbert (Gilbert *et al.* 1990) montrent que l'être humain commence par accepter l'information communiquée avant d'éventuellement la réviser ou la rejeter. On peut aussi penser aux travaux de Grice: la conclusion du schème argumentatif de l'appel à l'autorité fait émerger une explicitation de la maxime de qualité (Grice 1979) de toute proposition.

Avant cela, on peut voir dans les exemples ci-dessous, tirés de la presse, que la proposition «Donc c'est vrai»<sup>5</sup> est systématiquement ironique: l'énoncé «Donc c'est vrai» viole ostensiblement en effet une maxime de quantité, puisque dire que l'information X est vraie n'est en principe pas nécessaire. C'est précisément le caractère ostensif de cette violation de maxime qui fait émerger la lecture ironique du propos.

- (6) C'est écrit dans le journal, donc c'est vrai. (La Liberté, 21.12.2002)
- (7) «La dépêche l'a dit, donc c'est vrai!», résume un journaliste de LCI. (Le Monde Diplomatique, 1.06.2007)
- (8) Parce qu'ils vont coter les équipes de foot en Bourse, c'est mon mari qui me l'a dit, donc c'est vrai. (24 heures, 14.02.2006)
- (9) Le contrôle positif dont a été victime Hans Knauss est considéré comme «une immense surprise, certainement une erreur dans notre sport si beau et si propre» (c'est l'envoyé spécial de votre télévision qui le dit, donc c'est vrai). (Le Matin, 22.12.2004)

Ces exemples montrent par ricochet que l'apport informationnel de poser «X est vrai» peut être considéré comme si faible que l'explicitation de la conclusion n'a en principe aucun intérêt à être faite. D'une part, elle pourrait être perçue comme ironique, créant l'effet inverse de mise en doute des autorités citées (comme l'illustrent les exemples cités); d'autre part, la force de la stratégie argumentative liée à l'appel à l'autorité, elle, n'est pas diminuée par l'absence de conclusion implicite. C'est même tout le contraire, allons-

5 Ce corpus, non représentatif, présente quelques cas, assez rares pour plus de dix ans d'archives en ligne (base de données consultée: smd) d'occurrences de l'énoncé autonome «donc c'est vrai» - en éliminant les clivées «c'est vrai que».

nous arguer plus loin avant de prolonger notre réflexion à partir de toutes les conséquences de cette observation.

### 3.2 *Structure subordonnée de l'argument d'autorité*

Notons que cette double caractéristique de l'argument d'autorité le rend particulièrement difficile à illustrer à des fins pédagogiques. On peut en effet observer la difficulté qu'ont plusieurs manuels de l'argumentation à présenter des exemples canoniques d'argument d'autorité. A titre d'exemple, le manuel *Good Reasoning Matters!* (Groarke et Tindale 2004) introduit l'argument de recours à l'autorité en prenant appui sur un exemple fabriqué, dont on peut supposer qu'il est prototypique pour les étudiants visés par l'ouvrage:

- (10) «Comme le dit le chirurgien-chef (surgeon in chief), le tabagisme passif est mauvais pour la santé. Donc tu nuis à tes enfants lorsque tu fumes à la maison» (2004: 260, notre traduction).

L'exemple montre que le recours à l'autorité en question n'intervient pas dans l'enchaînement textuel marqué par une proposition A (le tabagisme passif est mauvais) donc C (tu nuis à tes enfants). L'autorité servant à légitimer la vérité de A *se greffe* sur l'enchaînement argumentatif explicite en A donc C. Autrement dit, l'autorité mentionnée vient renforcer le caractère indiscutable de la prémisse, mais, de prime abord, ne semble pas jouer un rôle dans l'enchaînement entre prémisse et conclusion explicites. Il en va de même pour l'exemple d'un manuel de Douglas Walton:

- (11) «Dr. Phil affirme que donner un pourboire diminue l'estime de soi pour celui qui le reçoit; donc, ne donne pas un pourboire à ce serveur» (2006: 87, notre traduction).

L'argumentation combine explicitement une prémisse imposée par le recours à une autorité et sa conséquence. Mais l'argument d'autorité fonctionne à l'arrière-plan, «se contentant» de légitimer la prémisse de manière subordonnée. Les spécialistes des structures de l'argumentation y verraient une argumentation complexe – une structure subordonnée (cf. Snoeck Henkemans 1997, Walton 1996, Herman ms., Herman et Micheli à paraître) – plutôt qu'une structure simple, présentant une forme d'emboîtement syllogistique que l'on pourrait représenter ainsi:

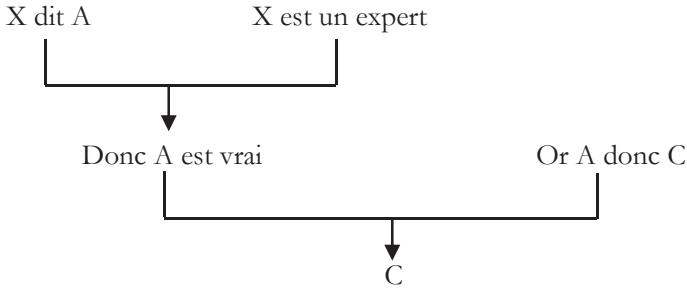


Figure 1. Structure complexe de l'argument d'autorité

En d'autres termes, alors que l'argumentation, selon une définition classique, met en scène une mécanique qui est constituée d'une conclusion justifiée par une prémisse, l'argumentation par l'autorité soutient le plus souvent une prémisse d'une de ces argumentations en laissant largement implicite tout le raisonnement qui légitime ladite autorité.<sup>6</sup> Dès lors, analyser l'argumentation qui fait recours à l'autorité selon un schéma syllogistique classique est si artificiel que plusieurs manuels montrent des exemples hybrides dans lesquels l'autorité joue un rôle d'appui à une argumentation, à l'instar de l'exemple suivant de Tindale pour lequel l'autorité paraît singulièrement peu saillante :

- (12) De récentes études scientifiques ont lié la consommation de l'Aspartame à des formes de cancer parmi des animaux de laboratoire. Étant donné la prévalence de l'Aspartame dans les boissons allégées, il devrait être retiré du marché. (2007: 133, notre trad.)

La structure particulière de cette argumentation, avec cette encapsulation de l'argument d'autorité proprement dit dans un enchaînement de surface en A donc C qui fonctionne avec un autre schème argumentatif, nous paraît avoir un effet très important: celui de faire passer la proposition P, garantie par

- 6 Il s'agit du cas de figure le plus fréquent. On peut imaginer un exemple du type non-subordonné: «Nous ne sommes pas allés sur la lune ; c'est l'ingénieur pour la fusée Apollo Bill Kaysing qui l'affirme». Cet exemple est réaliste pour une raison: l'ordre rétroactif (conclusion puis prémisse) évite la redondance et l'effet ironique de l'ordre progressif («Bill Kaysing dit qu'on n'est pas allés sur la Lune, donc c'est vrai»). Mais cet exemple se rencontre rarement, probablement parce qu'il nuit à l'efficacité du propos en exhibant une dépendance à la parole autoritaire qui est dès lors mise en débat alors que l'autorité est posée comme une évidence indiscutable dans les exemples subordonnés classiques.

l'autorité, comme acquise auprès de l'allocutaire; autrement dit, l'effet consiste à tenter d'*imposer la vérité ou l'indubitabilité a priori* de P vis-à-vis de celui qui constitue la cible de l'argumentation. Soyons bien clairs: il ne s'agit pas de considérer P comme indiscutable ou indubitable en soi – bien évidemment – mais de voir que la dynamique argumentative et interlocutive semble imposer P dans la situation d'énonciation comme une évidence censée être peu contestable. En somme, parce qu'on a affaire à une argumentation complexe dont l'autorité fonctionne comme une garantie en structure profonde, la contestation de P n'est pas prévue dans l'échange – même si elle reste toujours possible. A nos yeux, ce fonctionnement est très proche de la présupposition linguistique dans laquelle le locuteur impose à l'allocutaire la réalité d'une proposition. Dans l'exemple rabâché «Avez-vous arrêté de battre votre femme?», le fait de battre sa femme est considéré comme acquis, au même titre que, dans «la météo annonce du beau demain, donc prépare-toi à faire ton saut en parachute», le beau temps est présenté comme indubitable, certain ou admis. L'enchaînement par «Est-ce que tu crois tout ce que dit la météo?» est bien sûr possible, mais imprévu; il serait perçu comme une forme de retour en arrière sur un objet A qui était supposé être un objet d'accord préalable ou présenté comme tel (cf. Perelman), puisqu'il avait déjà permis son enchaînement par «donc C».<sup>7</sup>

De manière corollaire, cette dynamique argumentative met en exergue que la force de l'argument d'autorité consiste peut-être moins dans l'établissement de la vérité de P que dans la certitude de P imposée à autrui. En somme, P est considéré non seulement comme vrai, mais aussi comme indubitable a priori. Et ce, grâce à l'enchaînement en «donc» qui confirme la certitude de P. En effet, on peut comparer les enchaînements suivants:

- (13) Selon Météo Suisse, il fera beau et chaud demain. Mais mes rhumatismes prétendent le contraire, crois-moi.
- (14) Selon Météo Suisse, il fera beau et chaud demain. Préparez donc vos maillots de bain.

Ces deux exemples nous semblent montrer que la prédiction météorologique reste bien une prédiction – en ce sens, la vérité de ce qui est prédiqué sera

7 Pour un développement de ce point, voir de Saussure dans ce volume. Le mécanisme appelé de présupposition discursive et son accommodation propose un lien évident avec cette réflexion.



jugée au moment de sa supposée réalisation – mais aussi que sa certitude manifestée par un futur épistémique est prise en charge par le locuteur uniquement en (14) et est dès lors imposée à l'allocutaire. De ce fait, on peut dire que *l'argument d'autorité ne devient argument d'autorité qu'à partir de la prise en charge par le locuteur*, en l'occurrence à partir de l'enchaînement en DONC. Cette observation fait aussi écho à la notion d'autorité polyphonique d'Oswald Ducrot. Ce dernier montre que c'est l'enchaînement de P par Q qui donne à P sa légitimité justificatrice: «L'énonciateur de P joue le rôle d'une autorité en ce sens que son dire suffit à justifier L de devenir à son tour énonciateur de Q, en se fondant sur le fait que la vérité de P implique ou rend probable celle de Q» (Ducrot 1980: 157).

Cela nous conduit à considérer que la définition classique de l'argument d'autorité est incomplète si on ne mobilise pas l'ensemble de la structure complexe présentée plus haut, autrement dit, s'il n'y a pas confirmation que le locuteur fait comme s'il considérait que P est vrai. Notons que nous ne disons pas que le locuteur assume la vérité de P: il peut en effet, par exemple à des fins manipulatoires, considérer que P est douteux sinon faux, mais l'imposer comme vrai dans la dynamique argumentative par un enchaînement «E dit P donc Q». On peut donc dire que le locuteur paraît prendre en charge la certitude de P et impose cette certitude à l'allocutaire.

#### 4. Examen des caractéristiques essentielles

A ce stade, l'argument d'autorité suppose les caractéristiques suivantes – aux trois caractéristiques originelles avons-nous ajouté la quatrième:

1. L'auteur de l'appel à l'autorité donne comme argument en faveur d'une affirmation P le fait qu'elle émane d'une source tierce
2. Cette source est (présentée comme) une autorité
3. P est donc vrai (ou faux)
4. La certitude de P est imposée dans une argumentation généralement subordonnée confirmant que le locuteur prend en charge la vérité de P.

Nous allons dans la suite de ce texte interroger les quatre traits ci-dessus en commençant par le dernier pour ôter peu à peu ce qui fait la chair de l'argument d'autorité et mettre ainsi ce schème à l'épreuve.

#### 4.1 *L'absence d'argumentation subordonnée*

On peut trouver de nombreux cas dans lesquels l'argumentant cite une affirmation (15) d'une source autorisée (16) sans toutefois la faire subordonner par un enchaînement en DONC. En voici trois (voir aussi la note 7):

- (15) -Est-ce que tu penses arrêter de boire des boissons allégées?  
-Ben, mon médecin estime que l'aspartame est nocif pour la santé...
- (16) Les rendements financiers suffisent pour ne pas descendre ce paramètre à 6,4%, selon Michel Thétaz, le patron de la maison de gestion genevoise IAM. (Le Temps, 26.01.2010)
- (17) Pour Thierry Pech, directeur de la rédaction d'«Alternatives économiques», il est urgent d'augmenter la pression fiscale sur les plus aisés. (Le Monde, 10.11.2011)

Si le premier exemple est inventé, les deux autres sont des chapeaux d'articles journalistiques – présentant dès lors une forte autonomie. En l'absence de tout autre cotexte confirmatoire, il faut bien considérer que ces exemples certes répondent au prototype de l'argument d'autorité: un propos P, une source E du propos et un degré d'expertise affiché par une fonction sociale (médecin, patron de maison de gestion, directeur de magazine économique). Et pourtant, toujours en l'absence d'un cotexte qui conformerait une adhésion du locuteur au propos rapporté, rien ne permet de dire que les locuteurs dans les trois exemples considèrent que P est vrai.<sup>8</sup> Michel Charolles l'avait remarqué en 1987: «La valeur énonciative de p n'est jamais

8 Les deux exemples journalistiques sont intéressants dans la mesure où, d'une part, le/la journaliste ne montre pas explicitement son adhésion conformément à l'idéal d'objectivité qui prévaut dans cette pratique socio-discursive, mais, d'autre part, mettre en exergue une telle citation et un tel expert conduit à considérer que l'avis a des chances d'être aussi considéré comme vrai et pertinent par le locuteur journaliste. Il n'en reste pas moins que le contexte de controverse dans lequel P surgit peut être largement implicite alors que c'est lui qui déterminera si P est bien de nature argumentative et non descriptive.

prise en charge par celui qui énonce 'selon A, p', cette valeur est seulement indiquée ou, si l'on veut, montrée. Le L<sup>0</sup> n'assume pas l'acte signalé dans p, il le décrit comme pris en charge par l'énonciateur A» (1987: 251). Et Charolles de conclure avec pertinence: «[...] lorsqu'il est utilisé à des fins d'autorité, le tour en 'selon' (par exemple 'selon Aristote') produit d'avoir un effet de distanciation [...] qui doit être corrigé dans la suite par une marque d'adhésion explicite» (*ibid.*). On soulignera pourtant que l'adhésion est souvent sous-entendue par le contexte, même si elle n'est pas assumée. Dans l'exemple (15), le fait de ne pas faire suivre l'énoncé par une contestation du propos du médecin et de commencer celui-ci par une interjection montrant une forme de gêne personnelle semble préférentiellement indiquer que le locuteur croit que son médecin a raison, mais c'est un pari que doit faire l'allocutaire et sur lequel il peut hésiter en l'absence d'indice tangible confirmatoire.

Il n'en reste pas moins que l'absence d'adhésion marquée ne permet pas la pleine réalisation du programme de l'argument d'autorité. En effet, la définition de l'argument d'autorité donnée plus haut ne tient plus:

La prémisse majeure: la source E est experte dans le domaine S qui contient la proposition P; la prémisse mineure: E affirme que la proposition P est vraie ou fausse; la conclusion: P est vrai (ou faux).

Il faudrait restreindre la conclusion à: P est vrai (ou faux) pour E – ce qui est impossible puisqu'il s'agit de l'équivalent à la prémisse mineure. L'argument d'autorité doit dès lors inclure une donnée du type: «je crois la source E» pour faire admettre «P est vrai (ou faux)».

A cet égard, on voit par exemple que (18) montre un recours à plusieurs sources anonymes qui se recoupent pour attester d'un état du monde, mais l'usage du conditionnel dit journalistique ne permet pourtant pas d'établir le fait. En ce sens, le recours à l'autorité du témoignage constitue une bonne raison pour tenir P comme vrai, mais n'offre pas la garantie qui est proposée lorsque le locuteur prend en charge P.

- (18) Selon différentes sources et des propos de proches de victimes rapportés par le site de 20 minutes, la fusillade aurait fait plusieurs morts, peut-être trois. Il y aurait également des blessés, au nombre desquels des policiers et le forcené. (Le Matin, 3.01.2013)

En fin de compte, on voit que, sans la marque de l'adhésion du locuteur, il est difficile de parler d'argument d'autorité. On insistera donc sur l'importance donnée à cette adhésion dans la définition formulée par Christian Plantin: «Il y a argument d'autorité quand le Proposant donne pour argument en faveur d'une affirmation le fait qu'elle ait été énoncée par un locuteur particulier autorisé, *sur lequel il s'appuie ou derrière lequel il se réfugie*» (Plantin 1996: 88, nous soulignons).

#### 4.2 *P doit-il être uniquement vrai ou faux?*

On voit dans la définition de l'argument d'autorité une constante: P est vrai (ou faux). Or, on l'a souligné plus haut, «il fera beau demain» est un prédictif qui, en tant que tel, ne se juge pas en termes de vérité/fausseté mais en termes d'assomption ou non d'une croyance. L'enjeu, pour le locuteur, serait de faire adopter à l'allocutaire le point de vue d'un tiers comme suffisamment fondé – parce que le tiers est un expert – plus que de faire admettre la vérité ou non d'un fait.

Nous pensons que l'effet de l'argument d'autorité peut se décrire selon deux formes: l'une lorsque P est une croyance attachée à un énonciateur, autrement dit une opinion ou un point de vue, l'autre lorsque P est une assertion factuelle.<sup>9</sup> Dans le deuxième cas, le fait que ce soit un *argument* d'autorité pourrait être discutable dans la mesure où P n'est pas censé faire débat, mais assurer une représentation du monde. Dans les exemples suivants, on peut en effet se demander si on a affaire à un argument d'autorité en tant qu'argument:

(19) Selon le chef de la gare, ton train a cinq minutes de retard. Donc si tu cours, tu peux encore l'attraper.

9 On considérera ici que l'opinion ou la croyance se distingue de l'assertion factuelle par le fait que P constitue une évaluation ou une interprétation de la réalité par une source (singulière ou multiple) alors que l'assertion se présente comme la description d'une réalité. «Il doit aussi pleuvoir à Lausanne» est une opinion; «il pleut aussi à Lausanne» une assertion. «Marie est stupide» est une assertion, mais l'adjectif évaluatif nécessite une prise en charge par le locuteur de la responsabilité de l'énonciation qui est de l'ordre de l'opinion.

- (20) Je t'appelle parce que le jardinier a constaté que notre groseillier est mort. Veux-tu le remplacer par un autre?

Dans ces deux exemples, on a l'impression que les autorités citées le sont certes comme sources d'information fiables, mais qu'il n'y a pas d'effet ni d'enjeu rhétorique majeur, hormis d'augmenter la force de la vérité de P qui aurait déjà été exprimée par «Le train a cinq minutes de retard» ou «Le groseillier est mort». D'ailleurs, que veut dire «augmenter la force de la vérité»? Revenons sur ce point.

Comme on le sait, l'assertion est un acte de langage qui postule «P est vrai» (Searle et Vanderveken 1990). Dans (21), le nombre de manifestants est montré comme vrai par le fait même de l'assertion:

- (21) Il y avait plus de 2000 manifestants: on peut parler d'un beau succès d'estime.

Si on ajoute à cet exemple une source faisant autorité – comme dans (22), est-ce que P est «plus vrai»? Non, à l'évidence: la vérité n'est pas une notion graduelle.

- (22) Selon la police, il y avait plus de 2000 manifestants: on peut parler d'un beau succès d'estime.

On ne peut donc pas dire que P est «plus vrai» avec autorité que sans le recours à cette dernière. Dans le cas de l'assertion factuelle, la définition de l'argument d'autorité n'est, à notre idée, pas tenable: ce n'est pas parce qu'il y a autorité que la conclusion devient vraie. Du coup, même si les exemples (19-20) répondent aux exigences de l'argument d'autorité – ils seraient probablement traités comme tels par les spécialistes de l'argumentation – son rôle argumentatif ne nous paraît pas saillant par rapport à un rôle descriptif. Par ailleurs, l'idée d'adhésion que nous avons dégagée de l'argument d'autorité paraît absurde dans un tel contexte: on peut admettre ou non le retard du train ou la mort du groseillier, mais pas y adhérer.

Alors quel est l'intérêt du recours à l'autorité lorsque cette autorité asserte un état du monde? *Rendre P plus résistant à sa potentielle mise en doute* par la mention d'une source présentée comme fiable. Autrement dit, le recours à l'autorité imprime une force sur l'établissement du fait pour lequel l'autorité est citée non en tant qu'influence possible sur la croyance de l'allocutaire, mais en tant que source potentiellement fiable empêchant plus efficacement la mise en doute de la vérité de P que sans présence d'une source explicite.

Il ressort de ces observations qu'il y a, à nos yeux, un «véritable» argument d'autorité fondé sur P seulement si P exprime un point de vue attaché à un énonciateur dans la mesure où c'est précisément l'autorité qui va être utilisée pour tenter d'imposer une opinion sur une autre qui reste concevable et légitime, comme dans (23):

- (23) Pour Nicolas Tissot, la remontée du dollar face au franc devrait se poursuivre. «Avec le flux de mauvaises nouvelles en provenance de la zone euro, il n'y a pas de raison que l'euro ne baisse pas davantage», argumente le responsable des devises à la Banque Cantonale Vaudoise.

Notons que (23) tente d'imposer P non seulement par l'argument d'autorité, mais aussi par un argument lié à la baisse de l'euro qui montre bien que la croyance en P peut être sujette à caution et est *au cœur* du débat. En (19) ou (20), P est imposé comme vrai en guise de *préliminaire* au débat qui suit P: faut-il courir ou remplacer le groseillier par un autre?

On trouve dans les deux cas l'idée d'imposer P à l'autre. Mais, du coup, la dimension argumentative lorsque P est une assertion factuelle peut parfois être très peu saillante par rapport à une visée descriptive. Ainsi (24) est un exemple classique de deux sources d'information en désaccord sur les faits, exemple qui ne tente pas d'imposer le fait évoqué comme la réalité (d'autant moins qu'il y a contradiction dans les faits et biais possibles des sources d'information).

- (24) Alors que l'Assemblée examinera le texte à partir de mardi, les partisans du mariage gay étaient 125 000 dans les rues de Paris, selon la police, 400 000 selon les organisateurs. Une mobilisation moins importante que celle des opposants à la loi il y a quinze jours. (Le Temps, 28 janvier 2013, nous soulignons)

Cet exemple ne montre aucune volonté de faire admettre l'une ou l'autre réalité, mais de donner les deux sources pour que le lecteur puisse se faire une idée approximative de la mobilisation réelle. L'enjeu est donc de pouvoir décrire une réalité par le recours à des sources fiables.

Notons toutefois que l'idée d'imposer une réalité sur une autre est écartée par l'absence d'un enchaînement marquant l'adhésion à l'une ou l'autre de ces réalités – à l'inverse de (19) ou de (20). L'idée d'imposer P – que ceci soit une réalité objective ou une croyance – comme celle de montrer une forme d'adhésion du locuteur à P demeurent donc deux critères pour parler d'argument d'autorité, mais dans le cas de la croyance, la nature argumentative reste saillante, alors que dans le cas de l'assertion factuelle, la

nature descriptive de l'établissement du fait semble l'emporter sur l'idée d'influencer les croyances d'autrui.

En résumé, on pourrait schématiser ainsi notre réflexion:

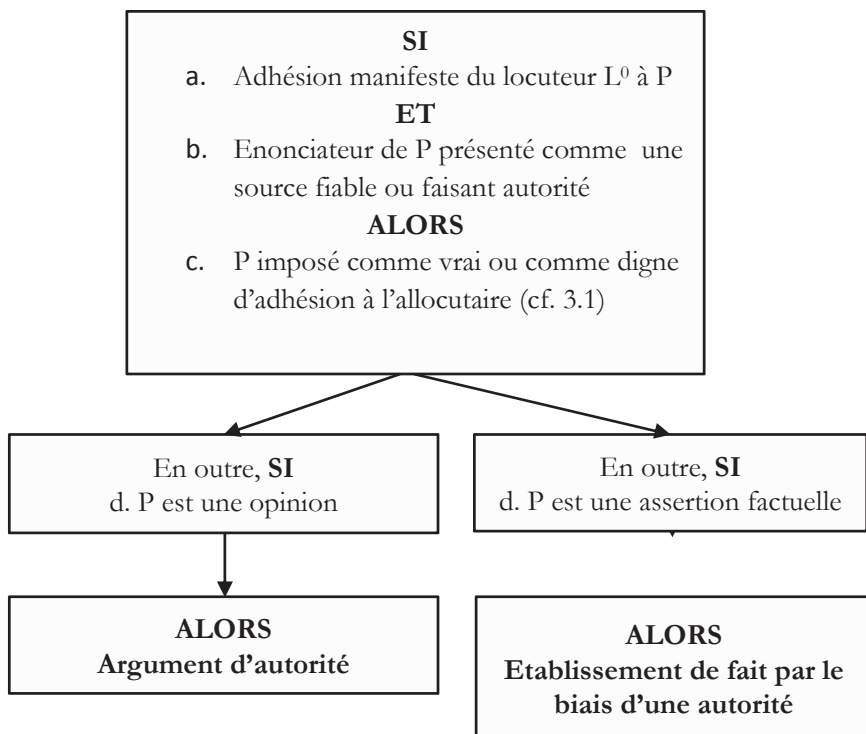


Figure 2. Conditions de l'argument d'autorité

#### 4.3 Quand la source n'est pas une autorité reconnue

(25) Notre voisin dit qu'il fera beau demain: peux-tu installer le grill et le parasol?

Est-ce que l'exemple (25) n'est pas un argument d'autorité? Il répond à toutes les caractéristiques vues jusqu'ici à l'exception d'une seule: on ne voit pas en quoi le voisin pourrait être une autorité reconnue en matière de météorologie. Et pourtant, le fait même de confirmer par Q (la demande d'installer le grill et le parasol) la croyance dans la prédiction P conduit à assumer, dans le

discours, une prémisse implicite selon laquelle le voisin est crédible lorsqu'il parle de météo. C'est à nouveau l'enchaînement qui non seulement peut confirmer une autorité, mais aussi l'instituer en tant que telle. On mesure par exemple que cela peut fonctionner à l'envers en instituant une contre-autorité:

(26) Si Jill dit que ce film est génial, ne va surtout pas le voir!

En la matière, nous sommes largement en accord avec la ligne Ducrot (1980), Perrin (2000): nul besoin d'une autorité prestigieuse pour un argument d'autorité:

Sans contester qu'un tel raisonnement [visant à établir la crédibilité de ce que l'on veut dire en l'imputant à quelqu'un de prestigieux] puisse être accessoirement mis en jeu dans le discours, et notamment dans certains cas particuliers d'argument d'autorité, je soutiendrai qu'il ne s'agit alors que d'une conséquence indirecte, et non du principe même d'un procédé par ailleurs tout à fait commun dans la conversation. Ni le prestige éventuellement imputable à la personne citée ni le raisonnement qui pourrait en découler – relativement à la vérité de son discours – ne sont généralement impliqués, en tant que présupposés d'aucune sorte, dans l'interprétation d'un argument d'autorité. (Perrin 2000: 82)

L'intérêt rhétorique majeur de l'argument d'autorité n'est pas prioritairement dans le prestige social de l'autorité citée, même si plus l'autorité citée est reconnue, plus l'impact de l'argument peut être fort a priori, mais dans le fait que le locuteur n'est plus seul à assumer le point de vue. Perrin, à la suite de Nølke, parle de dilution de la responsabilité énonciative du locuteur: à la fois «prise en charge par le locuteur et imputée à autrui» (2000: 89). Que l'autrui soit en effet un expert ou non n'a dès lors qu'une importance secondaire, puisque n'importe quel marquage évidentiel par ouï-dire peut fonctionner comme une marque d'argument d'autorité si le locuteur assimile le discours tiers au sien.<sup>10</sup>

Deux notes pour terminer ce point et pour illustrer des ressources manipulatoires de l'argument d'autorité: le trait d'évidentialité par ouï-dire implique que la source peut être vague ou non identifiable («il paraît que»)

10 Cette modalité a «pour fonction de préciser que ce qui est exprimé n'est pas le fait d'une création personnelle de la part du locuteur, mais tient à un discours-objet imputé – et emprunté – à autrui, c'est-à-dire à une forme de modalité allusive, à un argument d'autorité si l'on préfère» (Perrin 2000: 94).



voire inventée; le propos de la source peut être caricaturé: de la sorte, le sophisme dit de l'homme de paille (Oswald et Lewinski, ce volume) peut être considéré comme une sous-catégorie de l'argument d'autorité.

#### 4.4 *Quand la source n'est pas un tiers*

En extrayant de l'argument d'autorité l'idée d'une source tierce explicite, on peut se demander si on ne supprime pas un caractère essentiel pour l'identification du schème. Certains parlent pourtant d'autorité directe, mais «on remarquera [...] que la plupart des études sur l'argument d'autorité portent en fait sur sa forme indirecte, sa réalisation directe étant généralement mentionnée pour mémoire mais rarement développée» (Doury 1999: 4). Ce que nous voudrions donc observer, c'est un cas de figure où, au sein d'une argumentation, le locuteur avance (1) des faits ou des opinions imposées comme incontestables, (2) avec lesquels il manifeste son adhésion, mais (3) sans indice d'une source mentionnée ou existante.<sup>11</sup> Somme toute, il s'agit d'observer des assertions sous l'angle de l'autorité. En ce sens, nous aimerions faire écho à l'extension de l'argument d'autorité suggérée par Ducrot: «Une extension plus radicale encore de la notion d'autorité polyphonique consisterait maintenant – je me contente de signaler cette possibilité – à y introduire tous les enchaînements conclusifs, même ceux où l'asserteur de P semble identifié à L» (1980: 156).

Pour illustrer cette réflexion que nous ne pourrions qu'ébaucher ici, nous prendrons l'exemple suivant, un paragraphe de Guy Millière, dans un article en ligne paru sur Contre-feux (site n'existant plus à l'heure actuelle) dans laquelle ce spécialiste des Etats-Unis écrivait, le 4 septembre 2008, «Pourquoi je pense que Barack Obama ne sera pas élu». Cette prédiction – ratée – était dûment motivée par un certain nombre d'arguments, dont le premier était celui-ci:

- (27) (a) Le parti démocrate est, en fait, dans une crise profonde qu'il ne cesse de se dissimuler à lui-même. (b) Après avoir été l'incarnation de la gauche traditionnelle, plutôt social-démocrate, héritier de Franklin Roosevelt, (c) le parti s'est voulu, à la fin des années 1960, la 'nouvelle gauche' et de la contre-culture. (d) Cela lui a valu

11 Il s'agit d'un cas différent de l'auto-citation ou l'argument d'autorité auto-allusif étudié par Perrin (2000: 99 sqq.)

une seule victoire, celle de Jimmy Carter en 1976, et une série impressionnante de défaites, juste interrompues par les huit années de la présidence d'un Bill Clinton, (e) qui s'est présenté comme l'adepte d'un recentrement. (f) Le positionnement «nouvelle gauche» du parti démocrate attire à lui les minorités ethniques, les élites intellectuelles et professionnelles, ce qui peut rester de l'électorat social-démocrate, mais pas, ou fort peu l'immense classe moyenne (g) qui détermine les résultats finaux. (h) Obama incarne pleinement ce positionnement «nouvelle gauche», (i) qui lui a permis de remporter les élections primaires.

Cet exemple est saturé d'assertions qui prétendent décrire la réalité. Mais qu'est-ce qui garantit que l'on puisse se fier à P? Comme aucune source, aucun témoignage historique, aucune archive – bref, aucun élément de preuve ou *evidence* (en anglais) – ne vient témoigner de la vérité des propos (b), (c), (d), (e), (f), (g), (h) et (i), l'allocutaire n'a d'autre choix, en l'absence de connaissances historiques et politiques particulières, que de faire confiance à l'autorité du locuteur. Bien sûr, on peut imaginer que (d) est un fait partagé par une large communauté de lecteurs de cet article. Il est possible que plusieurs autres propositions confirment ce que certains allocutaires pensent ou savent déjà. Il n'en reste pas moins qu'en l'absence de ces connaissances-là, la seule ressource discursive donnée pour admettre P est la certitude exprimée par un locuteur. L'état supposé des connaissances doxiques est sujet à appréciation par le locuteur, et l'allocutaire qui admet déjà une proposition telle que (i) ne verra pas d'effet d'autorité si le locuteur assène (i) sans autre forme de preuve. Mais au-delà de ces variances situationnelles, les propos de Marc Angenot ne paraissent pas dénués de bon sens:

Nos croyances les moins douteuses résultent d'arguments d'autorité en dernière analyse: l'armistice de la Première Guerre mondiale a été signé le 11 novembre 1918; les Japonais ont attaqué Pearl Harbour le 7 décembre 1941; les mois d'hiver en Tasmanie sont juin, juillet et août; la Terre est ronde (Angenot 2008: 203).

En effet, si nous affirmons à un extraterrestre l'une des propositions de cette citation, il ne peut, en l'absence de tout autre élément de preuve, que se fonder sur l'autorité du locuteur pour admettre la proposition. Notons que, même si ce locuteur n'est pas une autorité, le principe de véracité de Reid montre qu'il a des chances d'admettre la proposition.

En passant aux opinions plutôt qu'aux faits, on peut dire que les généralisations du type «Les Français sont arrogants» fonctionnent selon un même procédé d'autorité imposée par le poids du présent, l'absence de toute modalisation qui donne à l'énoncé un aplomb difficilement contournable. Le

jugement ontologiquement subjectif y est présenté «comme-si» il était objectif – nous renvoyons aux travaux de Danblon (2002) sur cette question. Cette énonciation «délocutive» permet de présenter ce qui est dit «comme si le propos tenu n'était sous la responsabilité d'aucun des interlocuteurs en présence et ne dépendait que du seul point de vue d'une voix tierce, voix de la vérité» (Charaudeau 2005: 138).<sup>12</sup> Pour le dire avec les mots d'Ali Bouacha,

l'énoncé généralisant cumule cette propriété double et à première vue contradictoire d'exprimer un engagement et de permettre un désengagement. Engagement, puisque l'on cherche à imposer une proposition contre une autre proposition qui lui préexiste d'une manière ou d'une autre. Désengagement, puisque l'on fait comme si on n'énonçait pas son opinion personnelle, mais celle de tout le monde ou mieux encore une vérité fondée sur [...] ce qui 'oblige à penser d'accord' (Bachelard) (Ali Bouacha 1993: 285).

Mais, au-delà des multiples cas de figure qui peuvent se présenter, quels sont les éléments qui permettent de considérer qu'il y a un effet d'autorité du locuteur qui est imposé à l'allocutaire au sein d'une affirmation?

Pour répondre à cette question, il faut combiner une dimension textuelle et une dimension liée à la sémantique des modalités. La première condition pour considérer qu'une assertion joue sur des effets d'autorité directe du locuteur est que cette proposition ne soit pas étayée. Elle doit être imposée comme point de départ sans autre élément de preuve que l'énonciation même d'une proposition P supposée vraie:

1. P n'est ni justifié par une proposition Q ni validé par un élément de preuve externe – P se présente comme un point de départ d'une argumentation, point que l'allocutaire admet déjà ou doit admettre.

Pour ce qui concerne la question des modalités, nous nous fonderons sur la théorie de Laurent Gosselin (2010), lequel distingue dans son analytique des modalités plusieurs éléments constitutifs, dont I, l'instance de validation, et F, la force de la validation. Pour qu'il y ait un effet d'une assertion d'autorité, il faut que l'instance de validation ne soit pas la réalité comme dans «Cet arbre est un épicéa».<sup>13</sup> Autrement dit cela touche les modalités épistémiques («c'est une grande armoire»), appréciatives («c'est un joli film»), axiologiques («ce

12 Notons qu'à nos yeux, le locuteur incarne ou se fait le porte-parole de cette voix tierce.

13 Un test proposé par Gosselin pour repérer la modalité aléthique est l'insertion de «je trouve que» devant l'énoncé considéré. «Je trouve que cet arbre est un épicéa» transforme en épistémique l'énoncé aléthique alors que «je trouve que cet arbre est grand» garde la modalité épistémique présente dans «cet arbre est grand».

repas est immonde»). Nous considérons donc que la modalité aléthique («La fenêtre est ouverte») – qui «caractérise les jugements fondamentalement descriptifs (au sens où ils supposent que des faits leur préexistent, et où ils ont à en rendre compte), qui renvoient à une réalité existant en soi, indépendamment des jugements qui sont portés sur elle» (Gosselin 2010: 314) – ne présente aucun effet d'établissement d'un fait par l'autorité à la différence de la modalité épistémique qui désigne des «vérités subjectives» (*ibid.*: 325). Pour qu'il y ait effet d'autorité, il faut que la force de validation soit maximale, affichant une certitude absolue. L'énoncé «Obama incarne la position nouvelle gauche» est ainsi proposé comme certain et non probable ou contestable, ce qui couperait l'effet d'imposition d'une réalité à l'allocutaire. Enfin, même si la modalité est subjective, le jugement porté par l'énoncé ne doit jamais être présenté comme individuel («ce film est génial» crée un effet d'autorité en gommant le caractère individuel de l'appréciation, ce que ne fait pas «je trouve que ce film est génial» qui ne généralise pas l'appréciation à une collectivité). En somme, les deux conditions sont les suivantes:

2. P se présente linguistiquement comme une certitude, mais n'est pas aléthique.
3. P est subjectif, pris en charge par le locuteur, mais n'est pas individuel. L'énoncé est présenté comme validé par une subjectivité collective.<sup>14</sup>

Si ces trois conditions sont réunies, nous proposons de considérer que l'assertion P a un effet autoritaire mettant en jeu l'autorité du locuteur. Notons que nous ne parlons plus ici d'argument d'autorité, car on est trop éloigné du prototype décrit plus haut pour que cela reste pertinent.

14 Paradoxalement, «je suis sûr que Schlumpf est de gauche» répond aux critères (1) et (2), de sorte qu'on pourrait le considérer comme un argument d'autorité fondé sur l'autorité du locuteur, mais il ne répond pas au critère (3): c'est une vérité pour le locuteur, mais ce dernier n'impose pas la certitude de P à l'allocutaire – à l'inverse de «Schlumpf est de gauche». Dans la thèse de Thue Vold, elle affirme à juste titre: «Ainsi, *Ce traitement est sans risque* peut être perçu comme plus sûr que *Il est certain que ce traitement est sans risque*. Le fait que ce sont souvent les assertions 'neutres' qui sont perçues comme les plus sûres est dû à la polyphonie implicite dans les énoncés explicitement marqués comme certains. Par l'adjonction d'un modalisateur épistémique, le locuteur signale qu'il envisage la possibilité qu'il existe un doute ou un autre point de vue, et il doit donc souligner qu'à son avis ou en ce moment, il n'y a pas de doute» (Thue Vold 2008: 70).

On peut tisser des liens d'apparement entre l'assertion autoritaire et ce que Lescano (2009) appelle le «mode du trouvé» (2012) (anciennement «ton du monde» [Lescano 2009]), dans une perspective énonciative héritée d'Oswald Ducrot et de Marion Carel. Lescano met en effet en évidence que les énonciateurs, toujours ramenés à des questions de source des propos, gagnent à être plutôt vus comme des «tons» indiquant la «force de l'assertion» (2009: 47). Il distingue, à la suite de Carel (2011), trois modes énonciatifs: le mode du trouvé, le mode du vécu et le mode du conçu. Le «mode du trouvé»<sup>15</sup> correspond à la manière «factuelle de parler» (2012: 62): «On peut dire qu'un contenu mis en discours sur le mode du trouvé n'est pas censé faire l'objet d'une réaction négative. [...] L'allocutaire du mode du trouvé n'est pas construit comme 'réactant', mais comme simple 'auditeur'» (id: 63).

On voit l'intérêt de cette récente évolution de la polyphonie linguistique pour notre propos. Lescano met parfaitement en relief un des intérêts mode du trouvé qui est la question de l'indiscutabilité des propositions et relie cela, de manière implicite, à la question de l'autorité:<sup>16</sup> «l'énoncé qui parle sur un ton de Monde [mode du trouvé] se donne un auteur qui en est capable, un locuteur qui sait» (2009: 51).

Les implications de cette posture énonciative sur l'argumentation sont intéressantes, dans la mesure où l'apparence d'objectivité voire de factualité a pour effet de rendre inutile la convocation d'arguments servant à appuyer le propos (critère 1). L'assertion autoritaire devient telle par les effets de la déssubjectivisation ou par la transformation de propositions particulières en propositions universelles; mais cette façon d'imposer sa propre certitude conduit souvent à un effondrement des velléités d'argumentation. En effet, d'une part, le locuteur se donne les moyens par l'assertion autoritaire de ne plus avoir besoin d'arguments qui la légitimeraient, et, d'autre part, le destinataire n'est au minimum pas encouragé à proposer des contre-

15 Appelé ainsi parce que le «contenu est présenté comme ayant été 'trouvé tout fait' dans le monde» (Lescano 2012: 62).

16 Nous illustrons ici un rapprochement possible, mais ce n'est pas une équivalence. Ainsi, Lescano considère que «Je ne t'aime plus» pose son contenu sur le mode du trouvé. Pour notre part, nous limiterons l'assertion autoritaire à des constatifs (et non à des expressifs) dont le locuteur ne se présente pas comme la source énonciative de P. «Ce film est grandiose» est une assertion autoritaire du locuteur alors que «je trouve que ce film est grandiose» ne vise plus à imposer à l'allocutaire le contenu exprimé par P.

arguments – on pourrait même dire qu’il est incité à se taire. En effet lorsque l’on affirme: «la Suisse est l’un des pays les plus sûrs de la planète», on fait passer cet énoncé comme un point d’accord indubitable. Peut-être est-ce une raison d’ailleurs pour ne pas confondre *assertion autoritaire* et *argument* d’autorité, la première ayant pour effet de dissoudre l’argumentation en imposant des points de vue.

## 5 Conclusion

Le but de cette contribution était d’ouvrir la réflexion sur la question de l’argument d’autorité en partant des études classiques et normatives de l’argumentation et de leur définition pour la mettre à l’épreuve dans une perspective descriptive, fondée en sciences du langage et en tenant compte des effets cognitifs de cette ressource argumentative. Il s’agissait de bousculer l’argument d’autorité au point de conserver son effet principal – imposer la vérité de P à l’allocutaire – tout en allant jusqu’à éliminer ce qui est au centre de sa définition: une autorité externe. En distinguant certains cas de figure et en soulignant le rapprochement de formes linguistiques que l’on pourrait croire disparates, nous voulions donner une cohérence à une réflexion sur le fait de poser un contenu non seulement comme vrai, mais aussi comme indiscutable, quand bien même on ne fait usage d’aucune autre ressource que la confiance à autrui, et ce, tant du point de vue du locuteur («argument d’autorité») que du point de vue de l’allocutaire qui doit croire le locuteur («assertion autoritaire»). Il nous semble que cela ouvre de bonnes possibilités de dialogue entre les théories de l’argumentation, les sciences cognitives, la rhétorique, mais aussi, en particulier sur le dernier point, avec les approches de l’argumentation dans la langue.

Terminons par exemple avec une courte réflexion sur la forme verbale «on sait» en analysant deux exemples, parmi sept autres, tirés d’un même article scientifique (Guyot 2006).

- (28) **On sait**, depuis les récents travaux de Philippe Hamon ou Jean-Michel Adam et les polémiques par eux suscitées, que la description est loin d’être un objet simple à définir et à analyser.

- (29) **On sait** que ces deux genres ont pour point commun leur capacité à intégrer et à absorber des formes de discours extrêmement variées.

Ce qui nous a intéressé en particulier dans l'étude de cette forme langagière, c'est que la présence de la modalité épistémique est inutile sur le plan pragmatique. P et «on sait que P» étant redondants – puisque, si «on sait P», on peut se demander pourquoi devoir mentionner que l'on sache P –, la présence explicite du verbe savoir doit trouver sa fonction dans un effet de clan lié à l'autorité imposée. Le locuteur feint de rappeler ce que l'allocutaire est censé pourtant déjà connaître. Cela représente une manière de marquer le territoire – «nous partageons des connaissances communes» –, mais l'intérêt de la tournure ne se situe pas que là. D'abord, cela impose P comme indiscutable, car la prémisse mentionnée est présentée comme communément partagée, au sens le plus proche de ce que peut être une «donnée»; ensuite, cela pose le locuteur comme un expert du domaine, surtout si l'allocutaire du premier exemple n'a aucune idée des polémiques déclenchées à propos de la description; enfin, dans précisément le cas de figure où l'allocutaire «ne sait pas», cela force l'allocutaire soit à admettre par autorité le savoir en question, soit à abandonner le terrain.

En ce sens, la formule «on sait que P» paraît typique d'une assertion autoritaire, mais elle n'est certainement pas la seule. En outre, la possibilité d'un «on» multiple, communautaire, montre les liens possibles entre le fonctionnement d'une argumentation s'appuyant sur l'autorité et l'argumentation s'appuyant sur le nombre (*ad populum*). Elle nous paraît poser de manière cruciale la question des instances de validation des énoncés créant ou plutôt imposant des accords, des points de départ de l'argumentation que l'allocutaire, parfois, ne partage pas.

## Bibliographie

- Ali Bouacha, A. (1993): «La question générique: statut linguistique et enjeu discursif», in Moirand, S., Beacco, J.-C. & Ali Bouacha, A. (éds), *Parcours linguistiques de discours spécialisés*. Peter Lang, Berne, 279-289.
- Angenot, M. (2008): *Dialogues de sourds*. Mille et une nuits, Paris.
- Berrendonner, A. (1982): *Éléments de pragmatique linguistique*. Minuit, Paris.

- Bitzer, L. (1968): «The Rhetorical Situation», *Philosophy & Rhetoric* 1 (1), 3-17.
- Breton, Ph. (2005): *L'argumentation dans la communication*. La Découverte, Paris (4<sup>e</sup> éd.).
- Burge, T. (1993): «Content preservation», *Philosophical Review* 101, 457-88.
- Carel, M. (2011): «Attribution de point de vue et effacement du récit», in Corminboeuf, G. & Béguelin, M.-J. (éds), *Du système linguistique aux actions langagières*. De Boeck, Bruxelles 399-414.
- Charaudeau, P. (2005): «Quand l'argumentation n'est que visée persuasive. L'exemple du discours politique», in Burger, M. & Martel, G. (éds), *Argumentation et communication dans les médias*. Nota bene, Québec, 23-43.
- Charolles, M. (1987): «Spécificité et portée des prises en charge en 'selon A'», *Revue Européenne des Sciences Sociales*, XXV (77), 243-271.
- Cialdini, R. (2006): *Influence*. Collins Business Essentials, New York.
- Clément, F. (2006): *Les mécanismes de la crédulité*. Droz, Genève/Paris.
- Danblon, E. (2002): *Rhétorique et rationalité. Essai sur l'émergence de la critique et de la persuasion*. Éditions de l'Université de Bruxelles, Bruxelles.
- Doury, M. (1999): «El argumento de autoridad en situación: el caso del debate mediático sobre astrología», *Escritos* 17-18, 89-112. [version française online: <http://www.lcp.cnrs.fr/IMG/pdf/dou-99b.pdf>]
- Ducrot, O. (2000): *Le dire et le dit*. Minuit, Paris.
- Gilbert, D., Krull, D. S. & Malone, P. S. (1990): «Unbelieving the unbelievable: Some problems in the rejection of false information», *Journal of Personality and Social Psychology* 59, 601-613.
- Goodwin, J. (1998): «Forms of authority and the read *ad verecundiam*», *Argumentation* 12 (2), 267-280.
- Gosselin, L. (2010): *Les modalités en français. La validation des représentations*. Rodopi, Amsterdam.
- Grice, H. P. (1979): «Logique et conversation», *Communications* 30, 57-72.
- Groarke, L. & Tindale, C. (2004): *Good Reasoning Matters!* Oxford University Press, Oxford.
- Guyot, A. (2006): «Stylèmes et corpus génériques: un essai de confrontation au service de la stylistique des genres», *Corpus* 5, <http://corpus.revues.org/index472.html>.
- Herman, T. (2011): «Le courant du Critical Thinking et l'évidence des normes: réflexions pour une analyse critique de l'argumentation», in Micheli, R. (éd), *A contrario* 16, 41-62.
- (2010): «Either linked or convergent structures of argumentation? A false dilemma», communication à l'ISSA 2010, ms.
- Herman, T. & Micheli, R. (à paraître): «Quelques enjeux d'une théorie de l'argumentation à l'échelle du texte», in Adam, J.-M. (éd.), *Faire texte*.
- Kutrovatz, G. (2011): «Expert authority and *ad verecundiam* arguments», in Eemeren, F.H. van, Garssen, D., Godden, D., & Mitchell, G. (éds), *Proceedings of the seventh conference of the international society for the study of argumentation*. Rozenberg/Sic Sat, Amsterdam, CDRom, 1050-1061.
- Lehrer, K. (1989): *Thomas Reid*. Routledge, New York.
- Lescano, A. (2009): «Pour une étude du ton», *Langue française* 164, 45-60.
- (2012), «Le passé simple n'est jamais subjectif», *Tranel* 56, 61-75.
- Maillat, D. (2014): «Manipulation et cognition : un modèle pragmatique», in Herman, T. & Oswald, S. (éds), *Rhétorique et cognition*. Peter Lang, Berne.
- Millikan, R. G. (1987): *Language, Thought and Other Categories*. MIT Press, Cambridge, MA.



- Oswald, O. & Lewiński, M. (2014) : “Pragmatics, Cognitive Heuristics and the Straw Man Fallacy”, in Herman, T. & Oswald, S. (éds), *Rhétorique et cognition*. Peter Lang, Berne.
- Perrin, L. (2000): «L'argument d'autorité comme forme de modalité 'allusive' dans la conversation», in Martel, G. (éd.), *Autour de l'argumentation. Rationaliser l'expérience quotidienne*. Nota Bene, Québec, 81-106.
- Plantin, C. (1996): *L'argumentation*. Seuil, Paris.
- Plantin, C. (à paraître): *Dictionnaire de l'argumentation*. ENS éditions, Lyon.
- Reid, Th. (2000) [1764]: *An Inquiry into the Human Mind and the Principles of Common Sense*. Penn State Press, University Park, PA.
- Saussure, L. de (2014). «Présuppositions discursives, assertion d'arrière-plan et persuasion», in Herman, T. & Oswald, S. (éds), *Rhétorique et cognition*. Peter Lang, Berne.
- Searle, J.R. & Vanderveken, D. (1990): *Foundations of Illocutionary Logic*. Cambridge University Press, Cambridge.
- Snoeck Henkemans, A.F. (1997): *Analyzing Complex Argumentation: The reconstruction of multiple and coordinatively compound argumentation in a critical discussion*. Sic Sat, Amsterdam.
- Thue Vold, E. (2008): *Modalité épistémique et discours scientifique. Une étude contrastive des modalisateurs épistémiques dans des articles de recherche français, norvégiens et anglais, en linguistique et médecine*. Thèse, Université de Bergen.
- Wagemans, J.H.M. (2011): «The Assessment of Argumentation from Expert Opinion», *Argumentation* 25, 329-339.
- Walton, D. (1996): *Argument Structure. A Pragmatic Theory*. University of Toronto Press, Toronto.
- (1997): *Appeal to expert opinion*. Pennsylvania University Press, University Park, PA.
- (2006): *Fundamentals of Critical Argumentation*. Cambridge University Press, Cambridge.
- Walton, D., Reed, C. & Macagno, F. (2008): *Argumentation schemes*. Cambridge University Press, New York.
- Willard, Ch. (1990): «Authority», *Informal Logic* 12, 11-22.



# Argumentation from expert opinion in science journalism: The case of Eureka's Fight Club<sup>1</sup>

Sara GRECO MORASSO, USI, Lugano  
Carlo MORASSO, University College London

## 1. Introduction: a double-edged sword

In the modern and post-modern context, science pervades many aspects of the human society; scientists are involved in the public debate and it is impossible to do without experts in numerous institutional spheres as well as in private life. Science journalism is a promising context to study argumentation from expert opinion, because this type of journalism aims to transmit scientific results to the large public. Because of its nature and purpose, scientific journalism is a paradigmatic case of expert-layperson relationship, in which laypersons are bound to rely on the experts' opinion in order to get updates on current scientific developments (Caulfield 2004: 337). For this reason, we expect scientific journalism to be a context in which studying expert opinion is particularly important and promising.

In this context, the importance and the perils of authority in general and expert opinion in particular are felt both by scientists themselves and by the large public. Argumentation from expert opinion is feared because it is a double-edged sword: on the one hand, it may be accepted uncritically by the public for fear of going against the prestige of "science" (Walton 1997: 5-6; see also Fuller 2013: 40); this would be a typical application of the ad

1 We are most grateful to Sara Cigada for her helpful comments on an earlier version of this manuscript. Sara Greco Morasso would like to thank the Swiss National Science Foundation for a research fellowship (Grant PBTII-133595) which made it possible to write this contribution. Although this paper has been written in tight collaboration, it is possible to identify specific contributions by the two authors. Sara Greco Morasso has written sections 1, 3 and 5; Carlo Morasso has selected the case-study and collected the relevant data; he has written section 2 as well. Section 4 has been jointly written.

verecundiam fallacy (cf. van Eemeren & Grootendorst 1992: 135). On the other hand, the public may be wary of the asymmetry of information between them and scientists and be skeptical or overcritical towards expert opinion in general (see the results presented in Hargreave et al. 2003: 29-30).

Because of the latter risk, despite the fact that is possible to cite many debates in which scientists' initiatives have been well received by the public (as the danger from high blood pressure, and the negative health effects of smoking (Davis 2012), scientists feel they are not always understood (Chan 2005). Communication with the large public is a sensitive point, which is more and more topicalized in the scientific community. An editorial of the journal *Science* devoted to public engagement with science appeared as early as 2003 (Leshner 2003), while discussion is still ongoing (Sykes 2007, Bubela et al. 2009, Fuller 2013).

In this framework, approaches to communication with the public based on an indiscriminate use of argumentation from expert opinion are often seen as inadequate even by scientists themselves (Leshner 2003: 977). The awareness is increasing that laypersons should not only be instructed but respected as rational interlocutors. Therefore, the public should be involved in an argumentative partnership concerning the human, social and environmental implications of research, while it has been proven that massively relying on argumentation from authority increases the public-expert divide (Goodwin and Honeycutt 2009).

At the same time, on the public's side, as Walton (1997) argues, laypersons living in the contemporary society have to trust experts anyway for many issues in ordinary life. Yet they need to find a way to assess argumentation from expert opinion, in order not to be deceived.

From a cognitive viewpoint, our choice to focus on argumentation from expert opinion is driven by the fact that this type of argumentation presents what we might call a cognitive puzzle. On the one hand, in fact, relying on expert advice is to some extent unavoidable for a layperson to build well-informed knowledge in the context of cutting-edge scientific research. In this sense, expert opinion is part and parcel of a layperson's cognitive process of reasoning about science. On the other hand, however, misuses of expert opinion have been reported to block processes of argumentation in laypersons, particularly for fear of going against science (Walton 1997). Arguably, thus, if arguers make rhetorical misuses of appeals to expert

opinion, these latter might paradoxically become *arguments that block argumentation* –<sup>2</sup> therefore we have been speaking of a cognitive puzzle in this relation.

As a first approach to this complex topic, it is wise to start with a case-study in order to guarantee proximity to the reality we are analysing (Flyvbjerg 2001) and see how appeals to authority are used in a specific case of science journalism. In section 2, we will present the case-study that we are going to analyse, preceded by a short discussion of the communication context (Rigotti and Rocci 2006) in which it is placed. In section 3, we will discuss present studies on argumentation from expert opinion. Our theoretical framework is the Argumentum Model of Topics (AMT). Yet we will start from the informative contribution by Douglas Walton and extend it theoretically in two different senses. In section 4, on the basis of the theoretical instrumentation just discussed, we will analyse data and discuss one possible way in which the argument from expert opinion can go wrong. Section 5 briefly summarizes our findings and opens up new possible research paths.

## 2. Appeals to expert opinion in science journalism

In this section, we set out to briefly describe science journalism as the context of the use of appeals to authority. We will start from a general analysis of science journalism as a communication context (in the sense proposed by Rigotti and Rocci 2006) in which argumentation plays an important role (2.1); in section 2.2, we will describe the case-study we have selected and motivate our choice.

2 We believe it is noteworthy that J. H. Newman, in his *Essay in aid of a grammar of assent*, speaks of “*capricious ipse dixit of authority*” (our emphasis), having in mind misuses of authority which one needs to defend against in order to develop sound reasoning processes.

### *2.1 Science journalism as a context of argumentation*

Science journalism represents a peculiar context of argumentation as it is usually “spoilt” by two related problems which tend to hinder the development of an argumentative debate.

On the one hand, in fact, as science acts in an extremely regulated and defined way, where opinions are created, evaluated and falsified through a rigorous and laborious process, every counter-argument coming from other contexts is usually rejected a priori. There is, thus, a sort of non-permeability of the scientific debate to societal concerns (see Figure 1). In fact, scientific debate is closed to the general public and the arguments exchanged by scientists are not considered on the same level as, for example, economic arguments or arguments concerning potential social effects of scientific results. As Goodwin and Honeycutt (2009: 28) have shown, appeals to expert opinion in the public debate contribute to increase such non-permeability, because their use maintains “a sharp boundary between technical and public contexts”.

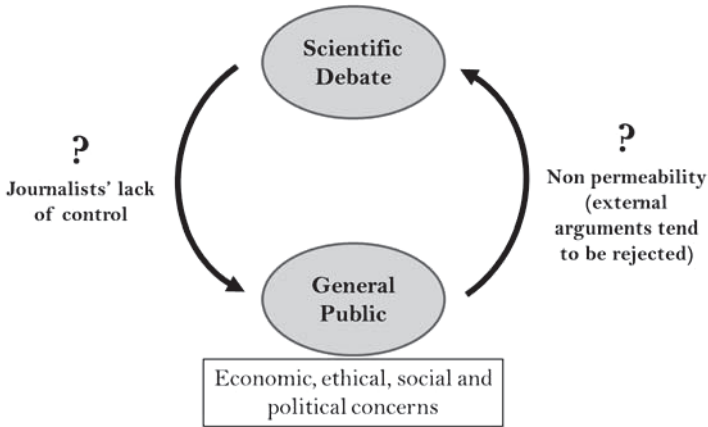


Figure 1. Sketched illustration of the problems affecting science journalism as context of argumentation.<sup>3</sup>

On the other hand, scientific debate is often non transparent for the general public. The supposedly informative mediation of science journalists, who should bridge scientists and the general public, is not always effective. The work of science journalists is to inform the general public of the advancement of science and to contribute to the divulgation of scientific opinions in the wider public. However, it is still debated whether science journalism is a canonical kind of journalism. While, for example, political journalists are an active part of the decision process on the topic they are referring, science journalists, because of their lack of time and because of the impossibility to be able to refer critically on every aspect of scientific knowledge, limit their action to the selection of topics and to reporting press releases by scientists. As a consequence, “science journalists are not players in the scientific process” (Murcott 2009: 1054). Besides, science journalists may lack the amount of knowledge or time which would be necessary in order to critically evaluate arguments put forwards by the experts; issues of accuracy may therefore arise (cf. Friedman, Dunwoody and Rogers 1986: 104ff).

3 The two main problems in the communication between scientists and the general public are indicated with a question mark.

Finally, a particular problem affecting science communication as a context of argumentation is related to the fact that there is a “cycle of hype” in science communication (Bubela et al. 2009, Caulfield 2009). This means that all the stakeholders involved in science face pressure to publish positive results. Scientists, research institutions and founding agencies face pressure to justify the money they spend while individual journalists need to define they work as newsworthy. As result of these factors, research projects tend to be described to the general public with a very positive attitude and are not critically evaluated (Caulfield 2009); by so doing, the commitment to reasonableness is jeopardized would be expected from an argumentative attitude (Rigotti and Greco Morasso 2009).

## 2.2 “*Fight club*”: a case study in science journalism

The case-study we have selected stands against this background as a “best practice” and an innovative one for the reasons we are going to discuss in this section. We have chosen to analyse the column *Fight Club* published within Eureka, a monthly supplement of the British newspaper The Times (<http://www.thetimes.co.uk>) launched in October 2009 (and then closed in October 2012). Building on The Times’ longstanding tradition in tackling science issues, Eureka’s specific goal was to cover a wide spectrum of topics which are currently debated at the scientific level and which significantly affect society. As the editors put it, “In launching Eureka, we recognise that many readers crave a better understanding of how science can transform our lives and our planet, which demands rigorous, engaging and inspirational reporting [...]. Its focus will be on the latest scientific developments”.

Latest scientific developments obviously tend to include controversial issues on which scientific debate is in progress. In particular, Eureka included a one-page double-column section, named “Fight club”, explicitly devoted to controversial issues, on which there is no unanimous scientific opinion. Our paper is based on the analysis of the Fight Club column; we take into account the fourteen issues published from October 2009 to November 2010. The types of discussion covered by Fight Club vary from human responsibility for global warming to commercial archaeology and from a reflection on human talent to the definition of autism (see Table 1).



<b>Issue</b>	<b>Page n.</b>	<b>Title</b>
October 2009	p. 13	Should conservation groups focus on cuddly, popular animals? David Nussbaum and Robert May decide the polar bears' fate
November 2009	p. 13	The origins of Man. Was it an accident that Neanderthals died out and humans survived? Two eminent scientists give their view
December 2009	p. 13	Is man largely responsible for global warming?
January 2010	p. 15	Is medical science doing enough to replace, reduce and refine animal experiments?
February 2010	p. 17	Is screen culture damaging our children's brain?
March 2010	p. 17	Is the science curriculum serving the needs of future scientists?
April 2010	p. 49	Should commercial archaeologists have a role in exploring our marine heritage?
May 2010	p. 49	Is talent taught rather than innate?
June 2010	p. 49	Should we cut our meat consumption to reduce carbon emissions?
July 2010	p. 49	Is autism a single condition?
August 2010	p. 49	Is pornography damaging to society?
September 2010	p. 49	Is over-cleanliness to blame for the rise in food allergies?
October 2010	p. 12	Are TV science stars more important than practicing scientists?
November 2010	p. 37	Should doping be allowed in sport?

*Table 1.* Eureka issues October 2009 – November 2010

Eureka's Fight Club might be considered a best practice about the use of expert opinion because it sets out to present a balanced view on a controversial issue. In fact, because two opposite expert opinions are proposed, scientific authority is challenged: the presence of two experts shows how scientific evidence on the issues considered is controversial.

Thus, *Fight Club* does not administer to the readers a monolithic scientific opinion. The intervention of scientific journalists is kept to a minimum, because the experts are directly drawn in almost all cases from the scientific world, in light of *Eureka's* goal is to present the public directly with the scientific dialogue. The intervention of scientific journalists, however, is still present in the selection of the experts.

### 3. Tools for evaluating appeals to expert opinion: the tradition of argumentation theory

#### 3.1 *Walton's critical questions*

Undoubtedly, we owe to Douglas Walton extensive and detailed studies on argumentation from expert opinion (cf. Walton 1997, 2006). On the backdrop of a careful and multidisciplinary review of several approaches to the problem of authority and expert opinion in various contexts, Walton famously defines six (now very well-known) critical questions to evaluate the argument from authority (Walton 1997: 223). In the following list, *E* indicates the Expert and *A* stands for Assertion:

- I. *Expertise* question: How credible is *E* as an expert source?
- II. *Field* question: Is *E* an expert in the field that *A* is in?
- III. *Opinion* question: What did *E* assert that implies *A*?
- IV. *Trustworthiness* question: Is *E* personally reliable as a source?
- V. *Consistency* question: Is *A* consistent with what other experts assert?
- VI. *Backup evidence* question: Is *A's* assertion based on evidence?

Walton explicitly says that these questions have been formulated with an eminently *pragmatic* aim in mind: i.e. that of helping us weight the strength of the arguments from expert opinion we are inevitably confronted with in our everyday life (financial advice, legal advice, science, medical consultations, and so on). Consistently with his purpose, Walton interprets the list of critical questions as a heuristic tool; he never claims to have analytically deduced *all*

possible critical questions. In his words: “It is my *recommendation* that the following six critical questions, matching the argumentation scheme for the argument from expert opinion, be called” (Walton 1997: 223, our emphasis).

In this sense, the selection of these six questions among others counts as a principle of economy and prudence (a *recommendation*). This leaves us with a theoretical problem. In fact, it is not clear whether Walton really wants to adopt the perspective of an analyst of argumentation; or if he is rather providing advice for laypersons to evaluate experts, as the term “recommendation” would seem to imply. The problem of aiming to a more consistent and comprehensive analysis of the argument from expert opinion from the theoretical point of view is therefore still open. In fact, if systematic criteria for the elicitation of critical questions are not provided, we will never be certain to have covered all possible critical questions for this argument scheme.<sup>4</sup>

In order to address this problem, we have two suggestions. First, we propose to “zoom out” and consider the broader picture of the human relationship in which argumentation from expert opinion may be used. Following Goodwin (2010), we take into consideration the expert-layperson human relationship as an *agency relationship*; this can help put many of the fallacious uses of the argument from expert opinion in the perspective of the human motivations behind the derailments of the arguers’ *strategic manoeuvring* (in terms of van Eemeren & Houtlosser 2002, van Eemeren 2010). We will discuss this in section 3.2.

Second, we suggest to “zoom in” on the relation between argumentation from expert opinion and the standpoint it supports. For this purpose, we adopt the Argumentum Model of Topics (Rigotti and Greco Morasso 2006, 2010) to analyse the inferential configuration of arguments and try to show how critical questions can be systematically associated to the different premises composing an argument from expert opinion (section 3.3). As it will become clear in section 4.3 especially, basing the analysis of argumentation from expert opinion on the AMT will allow discovering a new possible critical question.

4 A set of sub-questions is provided for each critical question. Also in this respect, however, no systematic and univocal criteria for the elicitation of sub-questions are made explicit.

### 3.2 *Agency theory and strategic manoeuvring*

Goodwin (2010) has proposed to extend Walton's theory on expert opinion by grounding his critical questions in a more general view of the expert-layperson relationship. Drawing on well-established acquisitions of economics and related disciplines, she regards the expert-layperson bound as an *agency relationship* (Eisenhardt 1989, Goodwin 2010: 136). In an agency relationship, "someone – the 'principal' – needs to retain someone else – an 'agent' – to do something she cannot or does not want to do for herself" (*ibid.*). Very often, the agent is an expert in some field in which the principal is a layperson; this latter needs to evaluate the expert's trustworthiness and reliability despite the asymmetry of information which intrinsically characterizes such type of relationship. In agency theory, as we might add following Palmieri (2010), not only the principal but also the agent is seen as a human being (rather than an instrument at the principal's disposal), who has his or her own desires and interests; and who, therefore, may want to engage in argumentation to persuade the principal about his or her objectives.

Notably, the agent's interests may under certain circumstances not be aligned with the principal's ones; in this perspective, agency explains why *conflicts of interests* may arise,<sup>5</sup> which kind of challenges these involve (cf. Goodwin 2010: 137), and what types of contracts are best suited to monitor agents in different possible scenarios.

Goodwin (2010: 138ff) suggests that reinterpreting Walton's critical questions in light of agency theory helps clarify why such critical questions are really *critical* and relevant to the expert-layperson relationship; it also helps

5 We have a prime example of how conflicts of interests may occur in the expert-layperson relationship in our corpus. In April 2010, the debate is devoted to this issue: "Should marine archaeologists have a role in exploring our marine heritage?" (Table 1). One of the two experts selected to answer to this question is Greg Stemm, the chief executive of Odyssey Marine Exploration which turns out to be a commercial archaeology company; and, should we trust what his counterpart suggests, one who is in huge debt. Not surprisingly, Mr. Stemm's standpoint is that "The private sector must be encouraged to help with the shipwreck resource". In such a case, when analysing argumentation from authority, the problem of conflicting interests should be taken into account, because it throws doubts on the expert's trustworthiness on this issue (Walton 1997: 227, Goodwin 2010: 138-139).

eliciting and using critical questions in a more *systematic* way. In particular, she suggests regrouping Walton's questions into two loose clusters, referring to two derailment of the expert-layperson relationship known as *moral hazard* and *adverse selection*. Moral hazard refers to lack of effort on the part of the agent, because the agent "may simply not put forward the agreed-upon effort" for a certain task (Eisenhardt 1989: 61). For example, if agents receive a fixed income which is independent from their results, they may decide not to work too intensively in order to reach the principal's goals.

Adverse selection is the misrepresentation of ability by the agent, who may claim to have certain skills when he or she is hired which the principal is not in the position to verify (*ibid.*). As Goodwin notices, regrouping Walton's questions in these two loose clusters emphasizes, "perhaps more than Walton did, that in judging whether to rely on what someone says, the layperson is assessing not the knowledge, but the *trustworthiness* of the purported expert" (Goodwin 2010: 141, our emphasis).

We would now like to bring Goodwin's suggestion a step further. The classification of Walton's questions under the labels of moral hazard and adverse selection allows focusing on the agent's and the principal's goals, thus linking more directly the possible "bad behavior" of the agent (expert) to his *strategic manoeuvring* in terms of van Eemeren & Houtlosser 2002 (see also van Eemeren 2010). We assume here the *extended* pragma-dialectical theory of argumentation (van Eemeren 2010), which claims that the arguer's dialectical aim to solve their difference of opinion on the merits by means of a critical discussion is always paired with a rhetorical goal. Each arguer (the expert being no exception) wants to win his cause; at the same time, he is committed to do it reasonably. Maintaining the balance between the commitment to reasonableness and the attempt at being effective means that the arguers have to *manoeuvre strategically in all moves that are carried out in an argumentative discussion* (van Eemeren & Houtlosser 2002). In particular, strategic manoeuvring manifests itself in the discourse through three aspects: "A particular choice made from the available *topical potential*, a particular way in which the opportunities for framing the addressee's perspective are used [*audience demand*], and a particular way in which presentational possibilities are exploited [*presentational devices*]" (van Eemeren & Houtlosser 2009: 6). Because the three aspects are always connected, van Eemeren (2010) proposes to picture them in a triangle to highlight their mutual interdependence.

Now, Goodwin's contribution, by emphasizing the problem of trustworthiness which, in turn, is based on the expert's own goals and desires, allows considering the expert's own contribution to dialogue with the layperson/principal as an *argumentative* contribution. The expert is an arguer and he/she has got his/her goals to pursue; therefore the principal must evaluate his contribution by carefully analysing and assessing the strength of the argument he proposes. Considering the agency relationship will thus help explain why the expert's strategic manoeuvring may derail, for example because the agent finds himself or herself in a situation of conflict of interest; and may use argumentation from expert opinion to cover a moral hazard or an adverse selection strategy.

### 3.3 *The Argumentum Model of Topics*

In section 3.2, we have reframed Walton's critical questions for argumentation from expert opinion in the broader picture of a human relationship, by using the concept of agency relationship. From the point of view of argumentation, this has permitted to connect possible manipulative uses of appeals to expert opinion to the expert's personal goal to win his cause; at this point, the pragma-dialectical notion of *strategic manoeuvring* has been introduced to explain derailments of appeals to expert opinion.

If we now zoom in on the details of the relation between a standpoint and an argument from expert opinion, we claim that a model such as the Argumentum Model of Topics, which allows distinguishing all implicit and explicit premises in an argument scheme (Rigotti and Greco Morasso 2009, 2010) helps reframe the critical questions more systematically and possibly add new ones, when necessary. It helps understand which premises are questioned by each single critical question.

The Argumentum Model of Topics (henceforth: AMT) was first proposed by Rigotti and Greco Morasso (2006) and then developed in a series of publications (cf. in particular Rigotti and Greco Morasso 2010). The AMT is compatible with the general framework of the pragma-dialectical approach to argumentation. Moreover, the AMT allows analysing the inferential configuration of arguments by distinguishing, on the one hand, premises of a *procedural* (formal) nature, directly depending on the locus, i.e.

the relation or principle of support connecting standpoint and argument. On the other hand, it accounts for *material* premises, connected to the speakers' cultural and personal experience.

At the core of the procedural component, we find the locus itself. i.e. the relation between the standpoint and the argumentation supporting it. When used to construct premises in real argumentations, loci give rise to a series of inferential connections called maxims. For example, in the case of the *locus from the final cause*, different maxims may be considered, such as "If a certain goal is to be pursued, it is reasonable to select a means to pursue it" or "If no means are available, the goal cannot be achieved" and some others (cf. Rigotti 2008). The maxims are then unfolded following logical forms, such as the modus ponens ("If no means are available, the goal cannot be achieved; and no means are available; therefore, the goal cannot be achieved"), modus tollens, or others.

Loci, maxims and logical forms give an account of the procedural component in argument schemes. Yet the analysis of argument schemes should not only account for logical principles; it should take into account real arguments used in real-life discussion. Therefore, in the AMT, it has been proposed to consider the intersection between procedural starting points and *material* starting points. Two material starting points have been distinguished: the *endoxon*, a general statement concerning cultural values or knowledge available to the arguers; and the *datum*, a premise of factual nature concerning the specific situation the arguers are considering.

In order to illustrate how the material and procedural components intersect in the AMT, we will briefly consider an example taken from Rigotti and Greco Morasso (2010: 498 ff) and refer to argumentation from *analogy*. It is based on a case of everyday life argumentation:

A: Should we travel by train or by car?

B: Remember the traffic jams on New Year's Eve? And today is our national holiday!

The argumentation here draws on the analogy between two celebrations, namely New Year's Eve and a national holiday, to conclude that travelling by car might not be wise on a holiday because of traffic. The graphical AMT representation of the analysis of this argumentation is shown in Figure 2.<sup>6</sup>

6 For a more detailed discussion, see Rigotti and Greco Morasso (2010).

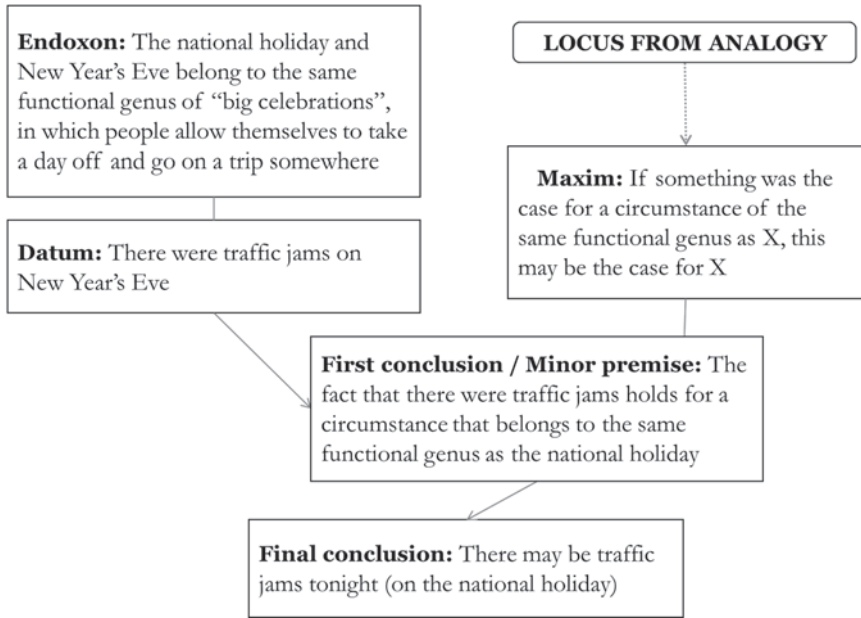


Figure 2. AMT representation of the New Year's Eve – National holiday argument from analogy.<sup>7</sup>

Each node of the diagram representing the AMT analysis is potentially subject to verification. Therefore, as Christopher Guerra (2008) has suggested, critical questions should be associated to each node. By this doing, the analyst of argumentation would guarantee completeness in the critical attitude towards an argument: in fact each of its premises, be it implicit or explicit, would be analytically put into scrutiny via the tool of critical questions. Within studies on the AMT model, this work is still in progress; and we wish to contribute to it with this paper. For the moment, Rigotti (2008) has given an example of how critical questions may be connected to

7 The procedural component is represented on the right of the quasi-Y structure of this graph. The material starting points (endoxon and datum) are on the left. The conjunction between endoxon and datum provides a "First conclusion" which is used as a minor premise in the procedural component. This First conclusion represents the intersection between the material and procedural components.



the nodes of the AMT model, by analysing in particular the procedural implicit premises of the means-end relation in pragmatic argumentation.

#### 4. Discussion

In what follows, we will discuss the results of our analysis, which we approach by trying to integrate the theoretical instruments presented in the preceding section. We will start with a general characterization of the column Fight Club in terms of the model of a critical discussion (section 4.1). In sections 4.2 and 4.3, we will turn to consider two main problems related to appeals to expert opinion which we have identified in our case study: vagueness in the material starting points and semantically unclearness in the procedural starting points respectively. We will see how these problems might be seen as derailments of the arguers' strategic manoeuvring.

4.1 Fight Club as setting the stage for two argumentative discussions

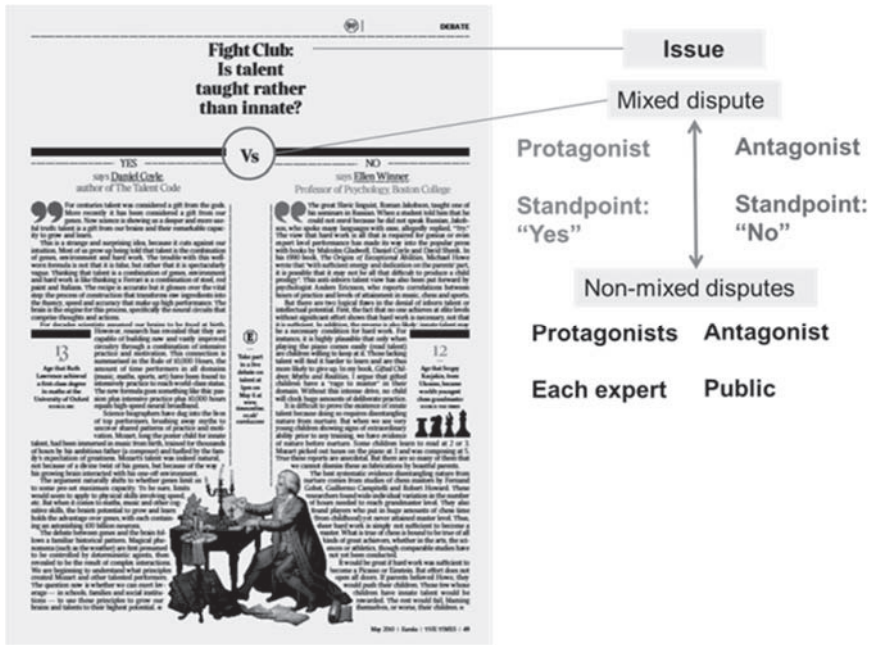


Figure 3. The Fight Club column: two argumentative discussions

At a general level, if we consider its *layout* from an argumentative point of view, the Fight Club column sets the stage for two argumentative discussions. Its structure is similar to what we find in talk-shows and other forms of media argumentation (cf. Goodwin and Honeycutt 2009). The issue of both discussions, namely the aspect concerned by the arguers' difference of opinion, is always formulated as a question addressed to the experts and to the public (Table 1). A first argumentative discussion occurs between the two experts, who play the role of Protagonist and Antagonist in a dispute which is pre-defined as *mixed* (van Eemeren & Grootendorst 1984) because the experts have opposite standpoints to defend. Such pre-defined standpoints correspond to the opposite answers to the question at issue ("YES" or "NO"). However, in several cases we found that the experts have a more nuanced view of their personal position and introduce different types of

epistemic modality; they formulate standpoints including linguistic mitigation devices such as “perhaps”, “maybe” and others (see Rocci 2005 for a discussion of this notion). One of them, M. Perkin, called to answer “yes” to the question whether over-cleanliness is to blame for the rise in food allergies, even jokes with the format of the Fight Club argumentative discussion in his opening statement: “With apologies to the larger advert, the answer is: probably” (September 2010).

The second argumentative discussion is between each expert and the public of laypersons who are the expected readers of *Eureka*. In this case, the dispute may be defined as *non-mixed*; in fact, the expert exclusively holds the burden of proof for his standpoint and he proves it in front of a public which is still to be persuaded and, as it is normal in mass-media communication, remains silent.

We face the expert-layperson asymmetry of knowledge only in the latter argumentative discussion. Therefore, in the following sections, we will focus our analysis of expert opinion on this discussion. However, this asymmetry is somehow mitigated by association to the former discussion in one and the same column: in fact, in the discussion with the public, the expert must take into account that his counterpart is a scientist as well. This guarantees a form of peer-review, as each expert knows that his or her arguments must be acceptable not only for the wider public but also for fellow researchers. Such form of indirect control guaranteed by the connection of two argumentative discussions is consistent with *Eureka*’s goal of presenting controversial scientific issues to the public.

#### *4.2 Vagueness in appeals to expert opinion*

In general, in our sample, appeals to expert opinion are not used to directly encourage the public to take the expert’s side. They are rather exploited to provide evidence and data which are not known to the public, because they are part of the scientific and not of the everyday life discourse. As a consequence of this preferential orientation in appealing to authority, those Fight Clubs whose basic evidence does not need much discussion do not present any argument from expert opinion. On the contrary, in those issues of *Eureka* in which the debate is about controversial evidence (e.g. is screen

culture damaging our children’s brain?), experts appeal to expert opinion frequently.

We advance the hypothesis that the reason for this preferential use of appeals to authority is to be found in the constraints imposed by the combination of the two argumentative discussions which the Fight Club experts are taking part in. In fact, while using authority in the form “believe me, do this” would be perceived as paternalistic by expert themselves (see section 2.1) and would not be acceptable in the argumentative discussion with the opposing expert, using the authority of a scientific study to corroborate evidence is a practice acceptable in the scientific community as well. This type of appeal to expert opinion, thus, works in both the discussions identified in section 4.1 because, if correctly used, it is acceptable even between experts.

Yet such appeals to expert opinion are not always used in a sound way. The major problem we have found when expert make use of appeals to authority is *vagueness* in the presentation of the source of authority – a problem already mentioned in Walton (1997: 139). The examples of vague appeals to expert opinion which we have found in our corpus are reported in Table 2.

	Month	Issue	Standpoint-expert	Appeal to expert opinion (text)
1.	Dec. 2009	Is man largely responsible for global warming?	YES, Chris Rapley, Director of the Science Museum and Professor of Climate Science at UCL	“The evidence is sufficiently compelling <b>for the science academies of America, Brazil, Canada, China, France, Germany, Italy, Japan, Russia and the United Kingdom</b> to conclude: “There is now strong evidence that significant global warming is occurring. It is likely that most of the warming in recent decades can be attributed to human activities”.

2.	May 2010	Is talent taught rather than innate?	YES, Daniel Coyle, author of <i>The Talent Code</i>	“However, <b>research</b> has revealed that they [our brains] are capable of building new and vastly improved circuitry through a combination of intensive practice and motivation”.
3.	June 2010	Should we cut our meat consumption to reduce carbon emissions?	NO, Maggie Gill, Scotland’s Chief Scientific Adviser on Rural affairs and the Environment	“Feeding the world while reducing greenhouse gas emissions is a key challenge. <b>Science tells us</b> that ploughing up vast tracts of grassland (on which sheep and cattle graze) would release millions of tonnes of carbon dioxide into the atmosphere and move arable agriculture on to land where yields are low. <b>Science</b> also tells us that grazing animals benefit biodiversity”.
4.	August 2010	Is pornography damaging to society?	YES, Mary Eberstadt, Research Fellow at the Hoover Institution in Stanford, California	“This year, <b>50 academics and other writers from left to right, including me</b> , collaborated on and co-signed <i>The Social Costs of Pornography: A statements of findings and recommendations</i> ”
5.	August 2010	Is pornography damaging to society?	YES, Mary Eberstadt, Research Fellow at the Hoover Institution in Stanford, California	“Then there is the harm that pornography obviously causes to children and adolescents. <b>One recent study</b> , funded by the US Congress, found that more than a third of teenagers report having been exposed to unwanted sexual content online. <b>Another reputable study</b> found that one in seven of 1’500 adolescents reported unwanted sexual solicitation online”.

6.	February 2010	Is screen culture damaging our children's brains?	YES, Baroness Greenfield, neuroscientist and director of the Institute for the future of the mind, University of Oxford	“If most of a young child’s actions take place on screen and so have no permanent consequences, it will prove a bad lesson when it comes to real life. <b>A recent study</b> found that obese people, for whom the sensual pleasure of eating trumps the consequences, are more reckless in performing tasks that involve an element of gambling. Could a daily life lived in the two dimensions of the screen be similarly predisposing the brain to a disregard for consequences?”
----	---------------	---	---	--

Table 2. List of the vague appeals to expert opinion found in our corpus

Vagueness in the indication of the source of authority makes it difficult to apply any of Walton’s critical questions. Let us take for example the Expertise question and all its sub-questions (Walton 1997: 223):

- How credible is E as an expert source?
- What is E’s name, job or official capacity, location, and employer?
- What degrees, professional qualifications or certification by licensing agencies does E hold?
- Can testimony of peer experts in the same field be given to support E’s competence?
- What is E’s record of experience, or other indications of practiced skill in S?
- What is E’s record of peer-reviewed publications or contributions to knowledge in S?

If our expert source is “research” or “science”, as it is in examples 2 and 3 in Table 2, asking the expertise sub-questions may be even meaningless. Similarly, the trustworthiness question is not applicable to “a reputable study” of which reputability (and, thus, trustworthiness) is predicated. The backup evidence question and the Consistency question do not test specific nodes of the single argumentation advanced by the expert, but intend to open

new paths (possible rebuttals) and invite to construct a broader context (co-text to be precise) which the expert's opinion can be weighed against. Clearly, not knowing exactly who the expert is makes it difficult to rebut; this problem even increases if the alleged expert is a monolithic entity such as "science". Therefore, at a first, rough analysis, we understand that vagueness poses problems to a sound use of expert opinion, thus making a layperson's reasoning processes difficult from a cognitive vantage point.

We now turn to analyse this type of vague appeal to authority with the help of the Argumentum Model of Topics in order to elicit implicit premises of a material and procedural nature and, thus, understand the level at which vagueness is working. Figure 4 represents a synthetic AMT representation of the appeals to expert opinion which are present in our corpus (cf. Table 2).<sup>8</sup>

8 We are aware that these arguments are different the one from the other and that a systematic analytical reconstruction of each and every one of them would bring to a series of different analyses and diagrams. Yet we would like to highlight what the different cases have in common – i.e. vague appeals to expert opinion – rather than their specific traits.

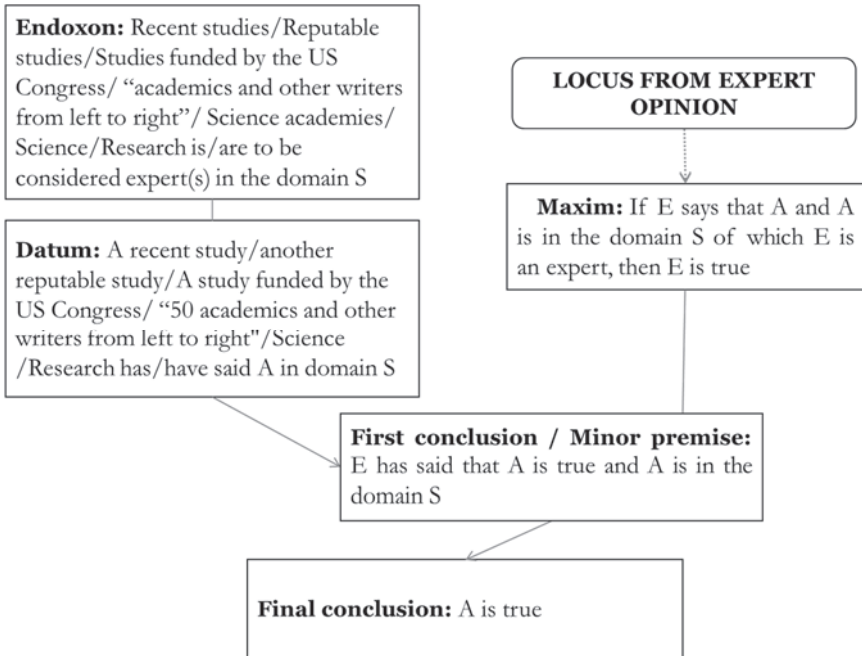


Figure 4. AMT adapted representation of the arguments from expert opinion featuring vagueness

As it appears in the AMT representation proposed in Figure 4, the main problem with this type of vague appeals to authority is to be found in the *material component* of the argument. In particular, the problem lies in the connection between the Datum, which is explicitly asserted in all of our examples, and the Endoxon, which is always implicit (cf. Table 2). Considering our corpus from the point of view of linguistics, we have identified two different types of vagueness.

A first type of vagueness is bound to the use of *abstract nouns*. The predicate “expert” in the endoxon is attributed to general abstract nouns, like “research has revealed...” or “science also tells us...” (Table 2, examples 2 and 3). By so doing, the resulting statements constituting the endoxa are incongruous from the semantic point of view because they violate a *categorial presupposition* (cf. Greco 2003): in fact, it is not science or research which can be experts; but single scientists or researchers.



\*Science is an expert in domain S

\*Research is an expert in domain S

The use of an abstract noun in this case is confusing. Abstract nouns, in fact, are nouns that refer to modes of being, state of affairs or events (Cigada 1999, 2004: 66). For example, “Linda’s relocation” refers to the event of Linda relocating somewhere; “The fall of the Roman Empire” refers to the Roman Empire falling. Similarly, “science”, used in this sentence, means “some people working in the scientific field” or “some scientists”; “research” means “some people doing research”. The use of an abstract noun, however, leaves the identity of such scientists or researchers very vague. This is certainly not accidental. On the opposite, it can be considered a specific strategy at the level of strategic manoeuvring of the expert contributing to Fight Club. In fact, leaving this information vague is a means to immunize one’s argument against specific critiques. Especially for a layperson, it is difficult to figure out whom the expert is referring to in order to verify the hold of his appeal to a (further) expert opinion. This is a derailment of strategic manoeuvring at the level of *presentational devices* in connection with the *topical potential*; in fact, the arguer pre-selects the evidence he or she needs, not giving any possibility to the antagonist (especially at the level of the argumentative discussion with the general public) to object or rebut, as no relevant counter-evidence can be introduced; and this represent a serious obstacle at the cognitive level for the development of a possible counter-argumentation. At the same time, the choice of these endoxa secures another advantage at the level of strategic manoeuvring, albeit an unsound one. In fact, evoking science or scientific research in general may trigger the functioning of the *ad verecundiam* fallacy, because a reader from the wider public may not dare go against the authority of “science” or “research”. Moreover, because the endoxon “science/research is an expert” is left implicit, it is even more difficult to challenge it for the readers. At the linguistic level, thus, this formulation of the argument exploits a mechanism of *presupposition accommodation* (Greco 2003) to immunize the topical selection of the expert’s strategic manoeuvring.

The second type of vagueness, which is to be found in examples 1 and 4-6 (Table 2), makes use of *concrete* nouns, i.e. nouns which refer to entities rather than to state of affairs. One might think that the use of concrete nouns is a way to eliminate vagueness; yet the entities that these nouns refer to are not

univocally identifiable (“one recent study”, “another reputable study”, “50 academics and other writers”); therefore, they remain vague, even though the violation of a categorial presupposition described above is not present in this case. Example 1 (“The science academies of America, Brazil, Canada, China, France, Germany, Italy, Japan, Russia and the United Kingdom”) seems to have a clearer reference but it might not be easy for a layperson to reconstruct what a science academy is and if the countries mentioned in this long list are really representative of scientific results. In all these cases, vagueness makes it almost impossible for a layperson to reconstruct the “identity” of the mentioned source of expertise; and not being able to identify the expert makes it difficult to propose contrasting evidence.<sup>9</sup>

Considering the vague appeals to expert opinion described above from the vantage point of agency relationship, the expert who is using them is trying and exploiting the asymmetry of knowledge which characterizes his/her relationship with the public; the expert is conveying the information that he or she is an indisputably reliable source and that the data he or she is presenting are unquestionable. This is to be interpreted as a case of *moral hazard* (cf. Goodwin 2010) which, at the argumentative level, as we have seen, is translated into a derailment of the expert’s strategic manoeuvring.

#### 4.3 *Towards the formulation of a new critical question*

Although the main problem related to appeals to expert opinion in our case-study was bound to the material starting points (see section 4.2), still there is one procedural aspect which needs to be mentioned. To some extent, in fact, problem of expert opinion in the modern times is bound to the maxim itself. In our analysis, as well as in Walton (1997, 2006), the maxim is formulated as follows: “If E says that A and A is in the domain S of which E is an expert, then E is true”. Such formulation, indeed, recalls the Medieval principle proposed by Peter from Spain: “*Unicuique experto in sua scientia credendum est*” (“any expert ought to be believed within his science/domain of

9 This strategy at the level of the expert’s strategic manoeuvring, which is present in both type of vagueness, could be read as a violation of the *freedom rule* of a critical discussion (van Eemeren & Houtlosser 2004), in the sense that immunizing one’s point of view goes against the dialectical context of a resolution of a difference of opinion.

expertise/field”, cf. Bochenski 1947). Yet a problem may arise about how to semantically define the notion of “field” of expertise, especially in the contemporary approach to science, which is highly interdisciplinary. As a matter of fact, at least four of the Fight Club issues we have been analysing show a problem in this relation; we will report them quoting from Table 1:

- October 2009: Should conservation groups focus on cuddly, popular animals?
- February 2010: Is screen culture damaging our children’s brain?
- April 2010: Should commercial archaeologists have a role in exploring our marine heritage?
- August 2010: Is pornography damaging to society?

It is quite clear that it is not straightforward to pinpoint at one single and well-defined field and discipline which these questions are referring to. Who is an expert on potential damages for children’s brain? Is that an exclusive matter for neuropsychologists? Are developmental psychologists relevant at all? Might scholars dealing with the notion of “screen culture” from the vantage point of media and communication studies be helpful in this relation? Similar problems arise with the other issues: respectively, the focus conservation groups should have, the role of commercial archaeologists in relation to marine heritage, and potential dangers caused by pornography to society. In all these cases, it is difficult to identify a clear-cut disciplinary field and the corresponding experts who could be used as sources for appeals to expert opinion. Eureka’s journalists make their choice and they have to restrict it to two experts only; of course this selection may be discussed because in some cases it might not be representative of the complexity of the selected field. Therefore, we would suggest to add a new critical question to Walton’s list, which could be formulated, as a working hypothesis, as: “Is the notion of ‘field’ applicable to this issue?”. As it clearly appears in this case, a semantic analysis of the maxim (and, in this case in particular, of the term “field”) of the type proposed in Rigotti (2008) and van Eemeren & Garssen (2009) may contribute to avoid manipulative uses of expert opinion.<sup>10</sup>

10 Moreover, as one of the reviewers pointed out, expert opinion is not to be considered equally valid in all issues discussed in Fight Club. In some cases, scientists are asked to comment on fact issues (e.g. *Is over-cleanliness to blame for the rise in food allergies?*), where

## 5. Conclusions

In this paper, we have analysed a case-study of argumentation in scientific journalism considering how appeals to expert opinion are used and if they are fallacious or not. The main focus of the paper was on the analysis of this empirical case. We have shown that a main source of potential manipulation in the use of expert opinion is vagueness in the indication of the experts who are being cited as sources of expert opinion. This problem is bound to the material component of argumentation, in particular to the endoxa which are evoked in argumentation. Another aspect which is a potential source of manipulation is bound to the semantic problem of how to define a “field” of expertise within the contemporary multidisciplinary approach to science. By analysing these problems with the AMT, we have shown that each node of the model, i.e. each implicit or explicit premise, may be subject to very specific critical questioning in order to test its soundness.

Despite its primarily empirical focus, however, this paper has equally contributed to the theoretical discussion of argumentation from expert opinion. Starting from Walton’s approach, we have extended it in two important senses. First, following Goodwin’s (2010) proposal, we have situated the problem of evaluating the expert’s trustworthiness in the broader framework of the agency relationship connecting experts and laypersons. This has made possible to interpret potential problems with the expert’s trustworthiness as derailment of his *strategic manoeuvring*, thus situating appeals to expert opinion in the properly argumentative framework of a critical discussion. In this relation, we have found evidence that derailments in strategic manoeuvring based on expert opinion represent attempts to block the development of a layperson’s cognitive processes of evaluation of expert argumentation, thus posing a serious challenge from the point of view of a sound development of reasoning in knowledge-oriented argumentation. In order to further pursue this line of research, it would be interesting to compare the results from social psychology reporting about non-

they are supposed to be knowledgeable. In several other cases, however, they are asked to discuss about values and policies, i.e. domains in which the use of appeals to expert opinion could be questionable.

argumentative teaching styles in education as potential barriers to learning processes (see for example Perret-Clermont, Carugati and Oates 2004).

Second, we have zoomed on the details of the relation between a standpoint and a supporting argument from expert opinion. For so doing, we have introduced the Argumentum Model of Topics (AMT) to show how the critical questions are connected to different kind of premises in the argument scheme. Relying on the AMT made it possible to give a more complete account of the problems related to appeals to expert opinion which are present in our case-study, by putting both procedural and material premises under scrutiny. This helps identify potential criticalities which go beyond the six critical questions proposed by Walton (1997), as shown in section 4.3.

## References

- Bochenski, I. M. (ed.) (1947): *Petri Hispani Summulae logicales*. Marietti, Torino.
- Bubela, T. *et al.* (2009): "Science communication reconsidered", *Nature Biotechnology* 27, 514-517.
- Caulfield, T. (2004): "Biotechnology and the popular press: hype and the selling of science", *Trends in Biotechnology* 22 (7), 337-339.
- Chan, K. M. A., Higgins, Paul & Porder, S. (2005): "Protecting science from abuse requires a broader form of outreach", *PLoS Biol* 3 (7), E218.
- Christopher Guerra, S. (2008): "Themen, Thesen und Argumente zur Position des Italienischen in der viersprachigen Schweiz", *Studies in Communication Sciences* 8 (1), 135-159.
- Cigada, S. (1999): *Nomi e cose. Aspetti semantici e pragmatici delle strutture nominali*. Milano, ISU. eBook available at: <http://www.educatt.it/libri/ebooks/C-00000208%20CIGADA%20-%20Nomi%20e%20cose.pdf> [March 15<sup>th</sup>, 2012].
- Cigada, S. (2004): "Nomi astratti che diventano nomi propri nel contesto dei messaggi pubblicitari", *Bulletin VALS-ASLA* 80, 63-70.
- Davis, K. C. *et al.* (2012): "Antismoking media campaign and smoking cessation outcomes, New York State, 2003-2009", *Prev. Chronic Dis.* 9, E40.
- Eemeren, F. H. van (2010): *Strategic maneuvering in argumentative discourse: Extending the pragma-dialectical theory of argumentation*. John Benjamins, Amsterdam/Philadelphia.
- Eemeren, F. H. van & Garssen, B. (2009): "The fallacies of composition and division revisited", *Cogency* 1 (1), 23-42.

- Eemeren, F. H. van & Grootendorst, R. (1992): *Argumentation, communication and fallacies: A pragma-dialectical perspective*. Lawrence Erlbaum, Hillsdale (NJ).
- Eemeren, F. H. van & Houtlosser, P. (2002): "Strategic maneuvering: Maintaining a delicate balance", in Eemeren, F. H. van & Houtlosser, P. (eds), *Dialectic and rhetoric: The warp and woof of argumentation analysis*. Kluwer, Dordrecht, 131-159.
- (2009): "Seizing the occasion: Parameters for analysing ways of strategic manoeuvring", in Eemeren, F. H. van & Garssen, B. (eds), *Pondering on problems of argumentation: Twenty essays on theoretical issues*. Springer, New York, 3-14.
- Eisenhardt, K. M. (1989): "Agency theory: An assessment and review", *The Academy of Management Review* 14 (1), 57-74.
- Flyvbjerg, B. (2001): *Making social science matters. Why social inquiry fails and how it can succeed again*. Cambridge University Press, Cambridge.
- Friedman, S. M., Dunwoody, S. & Rogers, C. L. (1986): *Scientists and journalists. Reporting science as news*. Free Press, New York.
- Fuller, S. (2013): "The public: Clients of science?", *Chemistry World* 10 (1), 40, [www.chemistryworld.org](http://www.chemistryworld.org) [January 12th, 2013].
- Goodwin, J. (2010): "Trust in experts as a principal-agent problem", in Reed, C. & Tindale, C. (eds), *Dialectics, dialogue and argumentation. An examination of Douglas Walton's theories of reasoning and argument*. College Publications, London, 133-143.
- Goodwin, J. & Honeycutt, L. (2009): "When science goes public: From technical arguments to appeals to authority", *Studies in Communication Sciences* 9 (2), 19-30.
- Greco, S. (2003): "When presupposing becomes dangerous. How the procedure of presuppositional accommodation can be exploited in manipulative discourses", *Studies in Communication Sciences* 3 (2), 217-234.
- Hargreaves, I. et al. (2003): "Toward a better map: Science, the public and the media", Economic and Social Research Council Report. [http://www.esrc.ac.uk/\\_images/towards\\_a\\_better\\_map\\_tcm8-13558.pdf](http://www.esrc.ac.uk/_images/towards_a_better_map_tcm8-13558.pdf) [January 11<sup>th</sup>, 2013].
- Leshner, A. I. (2003): "Public engagement with science", *Science* 299, 997.
- Murcott, T. (2009): "Science journalism: Toppling the priesthood", *Nature* 459, 1054-1055.
- Newman, J. H. (1870): *An Essay in Aid of a Grammar of Assent*. Burns, Oates & co, London.
- Palmieri, R. (2010): *The arguments of corporate directors in takeover bids*. Unpublished doctoral dissertation, USI-University of Lugano.
- Perret-Clermont, A.-N., Carugati, F. & Oates, J. (2004): "A socio-cognitive perspective on learning and cognitive development", in Oates, J. & Gravson, A. (eds), *Cognitive and language development in children*. The Open University & Blackwell Publishing, Walton Hall, Milton Keynes/Malden (MA), 305-332.
- Rigotti, E. (2008): "Locus a causa finali", in Gobber, G., Cantarini, S. Cigada, S., Gatti, M. C. & Gilardoni, S. (eds), *Proceedings of the LADA Workshop Word meaning in argumentative dialogue. Homage to Sorin Stati*. Special issue of *L'analisi linguistica e letteraria* 16 (2), 559-576.
- Rigotti, E. & Greco Morasso, S. (2006): "Topics: The argument generator", *Argumentum eLearning module*. [www.argumentum.ch](http://www.argumentum.ch) [March 2<sup>nd</sup>, 2012; restricted access].
- (2009): "Argumentation as object of interest and as social and cultural resource", in Müller Mirza, N. & Perret-Clermont, A.-N. (eds) *Argumentation and education: Theoretical foundations and practices*. Springer, New York, 9-66.

- Rigotti, E. & Greco Morasso, S. (2010): "Comparing the argumentum model of topics to other contemporary approaches to argument schemes: the procedural and material components", *Argumentation* 24 (4), 489-512.
- Rigotti, E. & Rocci, A. (2006): "Towards a definition of communication context. Foundations of an interdisciplinary approach to communication", in Colombetti, M. (ed.) *The Communication sciences as a multidisciplinary enterprise. Studies in Communication Sciences* 6 (2), 155–180.
- Rocci, A. (2005): *La modalità epistemica tra semantica e argomentazione*. ISU, Milano.
- Sykes, K. (2007): "The Quality of Public Dialogue", *Science* 318 (5855), 1349.
- Walton, D. (1997): *Appeal to expert opinion. Arguments from authority*. The University of Pennsylvania Press, University Park.
- (2006): "Examination dialogue: An argumentation framework for critically questioning an expert opinion", *Journal of Pragmatics* 38, 745-777.





# Two-sided rhetorical strategies in top management's letters to shareholders and stakeholders from corporate reports<sup>1</sup>

Ioana Agatha FILIMON, USI, Lugano

## 1. Two-sided communication in corporate reports

Corporate reports are communication tools by means of which listed companies account in front of their multiple stakeholders for the activity developed during a financial year, in order to build trust, facilitate (and influence) shareholders' investment decision, and legitimate themselves as responsible corporate citizens of the communities in which they operate. Although issued from past performance, corporate reports are future-oriented documents, aimed at influencing attitudes and decisions that are crucial for the future of a company. Both financial-economic annual reports (hereafter referred to as "annual reports") and corporate social responsibility reports (hereafter referred to as "CSR reports") share the same ontological roots – the practice of accounting for the (financial and non-financial) resources entrusted for administration – but stem from different rhetorical situations.

Introduced by Bitzer (1968), the concept of *rhetorical situation* designates

a complex of persons, events, objects and relations presenting an actual or potential exigence which can be completely or partially removed if discourse, introduced into situation, can so constrain human decision or action as to bring about the significant modification of the exigence. (1968: 6)

As literature suggests (Rutherford 2005: 352–353, drawing on Miller 1984, and Yates & Orlikowski 1992), corporate reporting genres can be defined by

1 This study was elaborated within the framework of the ProDoc research project "Endoxa and keywords in the pragmatics of argumentative discourse. The pragmatic functioning and the persuasive exploitation of keywords in corporate reporting", funded by the Swiss National Science Foundation (Grant PDFMP1\_124845/1) and coordinated by Andrea Rocci at Università della Svizzera Italiana (USI Lugano).

the three main constituents of the rhetorical situation that generates them: the *exigence* (a situational imperfection adjustable by means of discourse), the *audience*, and the situational *constraints* that influence the discursive action needed to modify the exigence.

Thus, annual reports stem from the necessity to bridge the information gap between managers (as agents) and shareholders (as principals), so as to enable the latter to make informed decisions about the company, and to evaluate managers' performance in administering the resources entrusted to them (Schoonraad, Grobler & Gouws 2005). Seeking to "construct a particular visibility and meaning, rather than revealing 'what was there'" (Hopwood 1996, in Stanton & Stanton 2002: 478), annual reports also have a pronounced persuasive character, their final goal being to attract investments in the company. Composed of both mandatory and discretionary disclosures, annual reports, as a whole, are obligatory for listed companies. The mandatory disclosures (mostly financial) must comply with reporting standards drawn in line with the accountability principle of a "true and fair view" (e.g. European Commission Directive 2006/46: 1; Financial Reporting Council (FRC), 2011) or "fair presentation" (e.g. US GAAP, in Sampsons 1985, Kirk 2006). This principle fosters an accurate presentation of a company's performance, and a realistic and prudent outlook on its strategic position (Kirk 2006, van Hulle 1993, FRC 2011), constraining the managers to address potential counter-attitudinal information in their reports. According to Sampson (1985: 1), "a fair presentation is *not* determined based on whether it is 'fair' to the company to disclose bad news"; a fair presentation should provide "an objective, neutral measure of performance, whether that performance is 'good' or 'bad'" (*ibid.*: 2). The discretionary disclosures, on the other hand, are not a direct subject of the reporting standards, having thus more space for rhetorical manoeuvres. Still, they need to be aligned with the mandatory disclosures in order to be credible (Rutherford 2005: 350).

In spite of the reporting standards in place, the credibility of the annual reports might still be questioned, since their impression management ends might prevail over the accountability ends (Stanton & Stanton 2002) and prevent a "fair" presentation of the corporate shortcomings. This is particularly true for the discretionary narratives, for which more rigorous provisions for a "true and fair" account have been suggested (Clatworthy & Jones 2006).

This credibility issue is even more pronounced in the case of corporate social responsibility (CSR) reports – voluntary disclosures emerged from the need of the companies to legitimate their activity in front of a wider range of stakeholders (like governments, communities, environmentalists, but also shareholders and employees) whose interests go beyond profit maximization, to social, environmental and ethical issues. Although CSR reports are an emerging genre, with evolving communicative conventions (Skulstad 2005), they usually adhere to reporting guidelines (e.g. UN Global Compact 2010; GRI Guidelines 2004) that, although less rigorous than the standards of the annual report, encourage compliance with the “fair view” principle. Thus, “a good ‘ethical’ report should be transparent and represent a genuine attempt to provide an account which covers negative, as well as positive aspects of all material impacts” (Adams 2004: 732). Yet, CSR reports are often accused of “window-dressing” (Schaltegger & Burritt 2006, Greenwood 2007), and “reluctance to report on negative impacts” (Adams 2004: 733, Stanton & Stanton 2002).

Additional constraints to address counter-attitudinal information in corporate reports may derive from the ontology of the firm. If companies are accused of wrongdoing, they have to protect their image in front of key stakeholders because they “cannot continue without those stakeholders” (Rowland & Jerome 2004: 197). And corporate communication research (e.g. Hearit 1994, 2006, Rowland & Jerome 2004) revealed a range of strategies of *corporate apologia* – “a broad term that means to respond to organizational criticism by offering a vigorous and compelling defence” (Hearit 2006: 4) – that are used by companies in order to repair or maintain a responsible image.

So, there are various accounting-informational or defensive rhetorical constraints that may lead managers to address counter-attitudinal information in corporate reports. Yet, their audience might still be sceptical about their intention to actually comply with these constraints, especially in the case of discretionary disclosures, which have more room for impression management maneuvers. Persuasion research (O’Keefe 1999) suggests that the choice of a *one-sided rhetorical strategy* (exclusively pro-attitudinal) or of a *two-sided rhetorical strategy* (that includes both arguments and counter-arguments to the advocated position) may have different consequences on the effectiveness of communication, in terms of persuasion and credibility – as further detailed in section 2. Therefore, considering the role of corporate reports in stakeholders’ decision making, and the impact of the latter on the

future of the companies, it seems worth investigating to what extent managers actually include counter-attitudinal information in their discretionary disclosures from annual and CSR reports, and with what possible rhetorical effects.

This paper will explore these aspects, narrowing the analytical focus on a sample of discretionary disclosures whose length and discursive structure are more appropriate for a detailed argumentative analysis: the introductory letters to shareholders (in annual reports) and stakeholders at large (in CSR reports). Arguably the most visible and influential narratives of corporate reports (Hyland 1998, Clarke & Murray 2000, Clatworthy & Jones 2006),<sup>2</sup> these letters are signed by a member of the top management – usually the CEO and/or the Chairman of the Board of Directors (named hereafter “managers”).<sup>3</sup> The explicit authorship increases the rhetorical complexity of the letters as compared to the “parent” reports since, according to Bitzer (1968: 8), an orator’s interests and communicative style can add to the constraints already in place in a rhetorical situation.

As parts of corporate reports, introductory letters should present a true and fair view of the corporate reality; still, as non-audited discretionary disclosures, they enjoy a relative freedom of topical choice (especially in CSR reports) that leaves room for rhetorical manoeuvring. That gives introductory letters “an enormous rhetorical importance” in establishing a relationship with the shareholders and building credibility, and also in conveying a positive image of the company (Hyland 1998: 224, Skulstad 2005). That is why, in spite of the “fairness” constraints entailed by the management’s stewardship towards shareholders (and, to a variable extent, towards stakeholders), introductory letters are perceived as a rather promotional genre, that might tend to present an overly positive corporate image (Hyland 1998: 224) – so favouring one-sided communication.

On the one hand, this tendency may derive from managements’ vested interest in attracting stakeholders’ financial and non-financial resources, on which the companies ontologically depend. On the other hand, managers’

2 Most studies focused on the introductory letters of the annual reports, and less on their CSR counterparts, but Skulstad (2005) pointed out a number of rhetorical similarities between the two types of letters.

3 Although the CEO and the Chairman have different roles in the management of the firm, they are both agents of the shareholders, on whose evaluation (based on the corporate performance) their position eventually depends. Besides, in many cases, the person who signed the letter was simultaneously both CEO and Chairman.

personal interests and communicative style may (intentionally or unintentionally) bias the information presented in the introductory letters – all the more as, although introductory letters represent the voice of the company, managers see them rather as representing them personally (Winsor 1993). Because their position depends on how their performance is evaluated by key stakeholders, managers may opportunistically try “to proactively shape shareholders’ and stakeholders’ perception of the organizational outcomes and events (*impression management*) and/or to retrospectively provide a self-serving account of events (*retrospective sense-making*)” (Merkl-Davies, Brennan & McLeay 2011: 320). Thus, managers may selectively report the events, so as to present themselves in a favourable, rather than in an unfavourable light (Clatworthy & Jones 2006). Also, they tend to attribute successful outcomes to internal corporate strengths (directly or indirectly pointing at management quality) and unsuccessful outcomes to environmental causes – a *hedonic bias* (Aerts 1994, in Stanton & Stanton 2002: 486) confirmed also by Bettman & Weitz (1983), Salancik & Meindl (1984), and Tsang (2002). In some circumstances, impression management motivation may conflict with professional credibility, managers finding themselves trapped in a *double-bind accountability dilemma* (Stapleton & Hargie 2011: 267), when all available explanations are avoidable due to their negative consequences at *ethos* level. For instance, if a manager admits that he has not foreseen a threat to the organization (e.g. a market fall), he indirectly admits his incompetence. But if he claims that he has foreseen the threat, but has not prevented the organizational loss, he’s likely to be seen as immoral. Therefore, he might try to avoid mentioning the event.

So, managers’ vested interest in conveying a positive corporate image (derived either from the ontology of the firm, or from their own opportunistic behaviour) may conflict with the rhetorical exigencies of this genre (that require the presentation of a “true and fair” view of the corporation) and influence the extent and the manner in which two-sided rhetorical strategies are actually employed in the introductory letters.

Therefore, this study conducted on two corpora of introductory letters extracted from annual, respectively, CSR reports will attempt to answer the following questions:

1. To what extent are two-sided rhetorical strategies actually used in the introductory letters of the two types of reports?

2. How do managers handle counter-attitudinal information in each category of introductory letters? Are there any corpus-specific patterns of two-sided argumentation?
  3. What possible impact may two-sided rhetorical strategies have on the communicative effectiveness of each category of introductory letters?
- 
2. Argumentation theory and persuasion research as analytical frameworks for the study of the two-sided rhetorical strategies of the corpus letters

The first two research questions of the study were answered by applying the analytical and interpretative tools of the theory of argumentation to the corporate reporting context. The argumentative reconstruction of the corpus letters – considered as an implicit argumentative exchange between companies (as protagonist) and their relevant publics (as implicit antagonist) – revealed the sidedness of managers' rhetorical strategies, identifying, for each corpus letter, the standpoints, the arguments and (whenever possible) the counterarguments, as well as the inferential and dialectical connections between arguments, counterarguments, and standpoints.

The theoretical framework for the argumentative reconstruction was the pragma-dialectical model of *critical discussion* (van Eemeren & Grootendorst 2004). This model assumes four (explicit or implicit) dialectical stages of argumentative interaction: a *confrontation stage* (where parties acknowledge their difference of opinion in relation to a specific issue); an *opening stage* (where parties – positioned as protagonist, respectively antagonist in relation to a standpoint – establish, from an assumed common ground, the starting points of their attempt to solve the difference of opinion); the *argumentation stage* (the argumentative exchange between the parties); and the *concluding stage* (where parties should agree whether the difference of opinion has been solved). According to this model, in each dialectical stage, arguers may appeal to various forms of *strategic maneuvering* (van Eemeren 2010) in order to resolve the difference of opinion in their favour, while maintaining the dialectical standards of reasonableness.

The pragma-dialectical model of critical discussion has multiple analytical advantages. It provides a pragmatic guideline for the detection and interpretation of the argumentatively relevant elements of the discourse, and normative standards for the evaluation of the effectiveness of the argumentative exchange (cf. van Eemeren & Grootendorst 2004: 58-59). As further detailed in section 3, this study also used this model as a pragmatic instrument for the generic, empirical description of each corpus of introductory letters, and as a ground for the integration of other complementary theoretical models of argumentative analysis – like the *Argumentum Model of Topics* (Rigotti & Greco Morasso 2010) – and of the theoretical and empirical hints provided by corporate communication research.

In order to estimate the impact of two-sided rhetorical strategies on the communicative effectiveness of the introductory letters – that is, on their capacity to fulfil their rhetorical objectives – this study combines two complementary theoretical frameworks. As presented in section 1, the introductory letters should be both credible – so as to build trust and maintain the relationship with the relevant stakeholders – and persuasive – in order to influence shareholders' and stakeholders' decision making in favour of the company. Thus, credibility could be considered a rhetorical purpose *in se* for these documents, though secondary to persuasiveness.

A first account about the impact of two-sided rhetorical strategies on the persuasiveness (and, to a certain extent, on the credibility) of the introductory letters comes from the normative criteria of evaluation of the reasonableness of argumentation, provided by the model of the critical discussion. These criteria entail that unreasonable arguments (based on false premises, logically unsound, or irrelevant in the reporting context) cannot effectively support a standpoint (e.g. a corporate stance). Analogously, if the counterarguments invoked by managers in their letters are (proved) unreasonable, they are ineffective in rejecting the corporate standpoints and should not damage the persuasiveness of the letters. It must be mentioned here that the pragma-dialectical notion of *effectiveness* is not completely synonymous with *persuasiveness*. While persuasiveness applies only to the arguments advanced in the argumentation stage of a critical discussion, effectiveness refers to arguers' efforts in all four dialectical stages in order to influence the resolution of the difference of opinion in their own favour (van Eemeren

2010: 39) – e.g. by strategically selecting or presenting the starting points in the opening stage.

The second evaluative account comes from persuasion research. While argumentation theory enables a theoretical evaluation of the two-sided rhetorical strategies, depending on the reasonableness of their constituent arguments, it lacks the empirical support for an estimation of the effect of these strategies – especially, of the variations in their argumentative patterns – on the actually perceived persuasiveness and credibility of the introductory letters. This gap can be filled by persuasion research that shows that, depending on the context and on their argumentative structure, two-sided rhetorical strategies may have different persuasive and credibility effects than those of one-sided rhetorical strategies. In a meta-analytical study about the message sidedness effects in advertising and nonadvertising contexts, O’Keefe (1999) observes that, in general, two-sided messages are perceived as more credible than one-sided messages, although not always as more persuasive.<sup>4</sup> Distinguishing between *refutational two-sided messages* that tried to refute somehow the opposing arguments, and *nonrefutational two-sided messages*, which acknowledged the opposing considerations without attempting to refute them, O’Keefe observed that in nonadvertising contexts, refutational two-sided messages appeared as more persuasive and more credible than one-sided messages. On the other hand, nonrefutational two-sided messages resulted significantly less persuasive than one-sided messages (presumably because they might have actually provided plausible counterarguments), and presented no credibility advantage (1999: 236–237). In contrast, in advertising contexts, two-sided messages did not result significantly more persuasive than one-sided messages – a persuasive advantage of the refutational two-sided messages being suspected, but not confirmed, due to insufficient available data. However, nonrefutational two-sided messages presented a credibility advantage over one-sided messages – an effect that has not been verified in the case of refutational two-sided messages (1999: 231, 237–238).

4 The terms ‘one-sided rhetorical strategy’ and ‘two-sided rhetorical strategy’ employed in this study are, in principle, synonymous with O’Keefe’s (1999) terms ‘one-sided message’, respectively ‘two-sided message’. The term ‘rhetorical strategy’ is preferred due to the heterogeneity of two-sided argumentation in the corpus documents, which often includes both refutational and nonrefutational argumentative moves that can be considered distinct two-sided rhetorical units in themselves.



Drawing on the Elaboration Likelihood Model (Petty & Cacioppo 1986), O'Keefe explains that an enhancement in message credibility may have heterogeneous effects on persuasion. On the one hand, credibility may function as a peripheral cue that favours persuasion, due to an enhancement of the general believability of the message. On the other hand, credibility might increase message scrutiny, so either favouring persuasion (if the quality of argumentation is good), or decreasing it (if the quality of argumentation is poor).

These findings suggest that the impact of the two-sided rhetorical strategies on the persuasiveness and the credibility of the introductory letters would be more comprehensively evaluated by a mixed approach that could combine the normative analytical rigour of the theory of argumentation with the explanatory and predictive strengths of empirical research on persuasion. This study makes one step in this direction, trying to extend the results of O'Keefe's (1999) meta-analysis to the context of corporate reporting, and makes a coarse estimation – at corpus level – of the impact of the two-sided rhetorical strategies on the persuasiveness and credibility of each category of introductory letters, depending on the relative proportion (and on the reasonableness) of refutational and non-refutational rhetorical moves they contain.

A precondition for the applicability of this mixed approach is to establish a correspondence between the types of contexts included in O'Keefe's (1999) meta-analysis, and the communicative context of the introductory letters. The main criterion used by O'Keefe in distinguishing advertising from nonadvertising contexts is message topic. He opposes topics that involve the advocacy for products, services, or sociopolitical issues, to socially-relevant topics that involve public questioning and debate (1999: 215, 240, Note 2). Apparently, the difference behind this topical demarcation resides in the alignment between parties' interests in the outcome of the communication, advertising contexts entailing a vested interest of the communicator in the success of the persuasive attempt, which cannot be a priori assumed in nonadvertising contexts.<sup>5</sup> This impression is reinforced by O'Keefe's

5 For example, producers ontologically depend (as a company) on the sale of the products they advertise. In contrast, participants to a debate aimed to resolve a social problem are more likely (although no necessarily) to be equally interested in any solution reasonably achieved through the debate.

hypothesis about the influence of the audience's initial expectations about the communicator on the credibility of two-sided messages (1999: 237-239),<sup>6</sup> which draws attention to the necessity of considering both audience characteristics and message topic in the estimation of message sidedness effects in a determined context.<sup>7</sup> Drawing on the empirically confirmed credibility advantage enjoyed by the communicators who advocate unexpected positions – like those that go against their apparent self-interest (Eagly & Chaiken 1975, Eagly, Wood & Chaiken 1978, Walster, Aronson & Abrahams 1966, and Wood & Eagly 1981 in O'Keefe 1999: 237) – O'Keefe hypothesizes that the enhanced credibility of nonrefutational two-sided messages in advertising contexts could be the result of the context-specific scepticism of the audience towards the intention of the communicator to present both sides of the issue. Moreover, contrary to the initial meta-analytical estimations, he hypothesizes that nonrefutational two-sided messages might have a similar credibility advantage in nonadvertising contexts, if the audience expects the communicator to intentionally avoid mentioning counter-attitudinal information. This was confirmed for the corporate context (Rogers & Stocken 2005), where investors often interpret bad news forecasts as more credible than good news forecasts, being aware of managements' institutional incentives to make earnings forecasts in a self-interested manner.

Thus, it can be assumed that, as an essentially accounting genre, introductory letters correspond to a nonadvertising context. As previously mentioned, managers are required to present a "true and fair view" of the corporate performance, and to account for the way they are fulfilling their duty towards the stakeholders whose interests they legally (or morally) represent. Yet, managers' vested interest in presenting an "embellished"

6 Audiences' initial expectations towards the communicator should not be confused with audiences' initial attitude towards the topic or the conclusion of the message. The consulted literature did not provide sufficient evidence for a reasonable assumption of audiences' initial attitude towards the corporate claims. If such an assumption would have been possible, it could have been connected to O'Keefe's meta-analytical findings in this respect (1999: 219, 229-230).

7 O'Keefe (1999) does not include audience's expectations within the context, which seems to be reduced in the meta-analysis to 'message topic' (e.g. 1999: 239, second paragraph). However, both Bitzer's (1968) model of rhetorical situation, and other models of communicative contexts (e.g. Rigotti & Rocci 2006) suggest that message topic and audience expectations are part of the same ontological frame.

image of the corporation is likely to elicit audience's scepticism about the impartiality of their report. In line with O'Keefe's (1999) meta-analytical findings and considerations about the message sidedness effects, it can be argued that in the communicative context of the introductory letters, refutational two-sided rhetorical moves might present a persuasive advantage, while nonrefutational two-sided rhetorical moves might have an opposite effect. *Ex hypothesi*, considering the audience's background expectations towards managers' impression management motivation, it could be hypothesized that non-refutational rhetorical moves might improve the credibility of the introductory letters – as previously argued, a rhetorical purpose *in se* for this genre – but they would benefit the overall communicative effectiveness of the letters only to the extent that they do not endanger the managers' perceived ability to fulfil their duty towards the relevant stakeholders.

### 3. Corpus description and methodological approach

This study was conducted on a corpus of 103 introductory letters (and equivalent interviews with top managers) extracted from 27 annual reports and 47 CSR reports, belonging to 22 listed companies, heterogeneous in terms of type of industry and country of origin. The reports covered the financial year 2007 and were published in 2008, in English, on the internet. One of the companies had 14 CSR reports and 23 corresponding letters, out of which 7 were identical. After the elimination of 6 identical letters, the final corpus of the study consisted of 44 introductory messages from annual reports, and 53 introductory messages from CSR reports.

The first methodological step was to identify, by coarse argumentative reconstruction, those introductory letters that used a *two-sided rhetorical strategy*. Then, these letters were subject to a more detailed argumentative reconstruction, in order to isolate potential genre-specific argumentative patterns in the treatment of their counter-attitudinal information. These first analytical steps were carried out in line with the pragma-dialectical model of critical discussion (Snoeck Henkemans 1997, van Eemeren 2010) and the

*Argumentum Model of Topics* (AMT) (Rigotti 2006, Rigotti & Greco Morasso 2010).

To guide and, then, synthesize the reconstruction and the evaluation of two-sided rhetorical strategies, this study also used the critical discussion as a pragmatic instrument for the generic empirical description of the two corpora of introductory letters. The basic configuration of the resulting *generic critical discussions* emerged from the rhetorical situation of each category of letters, and from the preliminary coarse argumentative reconstruction made in the first analytical step. Thus, the assumed generic protagonist of the introductory letters was the company, represented by the manager(s) that signed the letter. The generic antagonists were the (usually) explicit addressees of each type of introductory letter: the shareholders, in annual reports, and stakeholders, in CSR reports. Considering the persuasive ends of the annual reports, the generic issue of the letters to shareholders was whether the company was worth investing in, and the assumed pragmatic (and usually, implicit) generic standpoint was “You should (continue to) invest in this company”.<sup>8</sup> The generic issue of the letters to stakeholders was whether the company was behaving socially responsibly, and the assumed descriptive generic standpoint was: “This company behaves socially responsibly”.<sup>9</sup>

The AMT was used for the recovery of the starting points advanced in the opening stage of the two generic critical discussions, and the analysis of the main argumentative patterns employed in the argumentation stage, according to Rigotti’s (2006) *taxonomy of loci*.<sup>10</sup> Particularly relevant was the concept of *endoxa* – context-bound material starting points consisting in

8 The concept of *generic standpoint* for the argumentative analysis of a large corpus of documents was previously used in Filimon (2011).

9 Although the heterogeneity of the stakeholders and of their interests may generate distinct sub-discussions within the same introductory letter (between the company and various categories of stakeholders) these sub-discussions are eventually aimed to defend the social responsibility of the company as a whole. Therefore, this study assumed only one generic antagonist (stakeholders, at large) and one generic issue for the CSR letters.

10 *Argumentative loci* are indicated in the argumentative reconstruction by the *maxims* that underlie the argumentation schemes. *Maxims* are implicit procedural starting points consisting in the inferential implications entailed by the ontological relations on which the argumentative reasoning is based (i.e., *loci*) (Rigotti & Greco Morasso 2010: 495). Examples of *loci* are the ontological relations between cause and effect, whole and parts, etc. An example of *maxim* is “if the cause is the case, the effect is too” (*ibid*: 495).

explicit or implicit propositions that, being already accepted by the relevant audience, are used as major premises in argument schemes (Rigotti 2006: 527). If falsely assumed as already accepted by the audience, *endoxa* can be a source of fallacy.

Based on the results of the argumentative reconstruction, the third methodological step was to outline a *rhetorical profile* of each corpus of letters, based on the discrete *two-sided rhetorical moves* identified in the argumentation stage of the generic critical discussions (cf. Tables no.1 and no. 2). The two-sided rhetorical moves were classified according to two dimensions. The first dimension followed O'Keefe's (1999) demarcation between refutational and nonrefutational two-sided messages, superposed onto Snoeck Henkemans' (1995) distinction among refutations, concessions, and mere acknowledgements of counterarguments. Thus, a *two-sided refutational move* consisted of one standpoint advanced by the manager (either the main standpoint of a letter, or a sub-standpoint from the argumentative structure of that letter), a counterargument to this standpoint (e.g. a criticism or other negative aspect casting a doubt on the corporate performance), and the argument(s) by means of which the manager attempted to refute that counterargument. *Direct refutations* consisted of a *refutation proper*, i.e., a denial of the truth-value of the propositional content of the counterargument. *Indirect refutations* did not deny the truth-value of the propositional content of a counterargument, but denied (or attempted to diminish) its relevance – corresponding, thus, to Snoeck Henkemans' concept of *concession*. Analogously, a *two-sided nonrefutational move* consisted of one standpoint advanced by the manager and a counterargument to this standpoint; here, the counterargument was acknowledged, but not refuted in any way in the text.

For a more accurate demarcation of two-sided rhetorical moves, considering also that managers see introductory letters as representing them personally (Winsor 1993), this study observed also Schütz' (1998) taxonomy of defensive styles of self-presentation. This taxonomy includes: the *denial* of the negative facts; the recognition of the facts, but not of their relevance or connotation (by *reframing* or *dissociation*); the recognition of the facts, accompanied by *justifications* or *excuses*; and the full recognition of responsibility for the negative facts accompanied by *apologies* and some sort of *remediation* (Schütz 1998: 619). Considering the factual exigencies of the reporting genre, for analytical reasons, both apologetic elements and

justifications were included in the same two-sided nonrefutational rhetorical move as the negative event they accompanied – without ignoring, though, their potentially different consequences for managements' image.

The second dimension of classification of the two-sided rhetorical moves was the *time* of occurrence of the negative event referred in the counterargument. Some moves were related to past corporate events, being included in managers' strategies of retrospective sense-making. Other two-sided rhetorical moves addressed future risks for the corporate performance, while a third category of moves referred to problems affecting the companies on a continuous basis. It is true that *argumentation* intervenes only when the truth-value of a certain claim is under doubt; hence, past results – which are already a fact and whose existence cannot be put under doubt – can only be subject of *explanations*.<sup>11</sup> But when managers explain past results, they attribute them causes, which come either from inside the organization (*internal attributions*) or from outside of the organization (*external attributions*). These causal attributions are an indicator of the extent to which companies assume the responsibility for the reported results; therefore, they can be transformed in arguments for the ability of a company (as agent) to produce the effects expected by the stakeholders (like profits, ethical behaviour, etc.) – in other words, in *arguments from the efficient cause*.<sup>12</sup> Also, managers may use argumentation in order to diminish the relevance of past negative events in relation to the overall corporate performance (e.g. by presenting a minor loss as a means for the achievement of a major good).

The fourth methodological step consisted in the coarse estimation – at corpus level – of the possible impact of two-sided rhetorical moves on the communicative effectiveness of the corpus letters, based on the rhetorical profiles outlined in the precedent step. As presented in section 2, this estimation was made in line with O'Keefe's (1999) meta-analytical observations and hypotheses about message sidedness effects, taking into account the nonadvertising topic of the introductory letters, and the potential initial scepticism of the audience towards managers' intention to present both sides of the corporate activity.

11 The distinction between *argumentation* and *explanation* was made according to Walton (2006).

12 In general, in the corpus letters, managers' attributions referred to the responsibility of the corporation; only exceptionally (e.g. Chairman's Letter, in *BP Annual Report*) did they refer to the personal responsibility of a member of the management.

This estimation was then confronted to the argumentative analysis of the most representative two-sided rhetorical patterns identified in the two corpora of introductory letters.

#### 4. Results and discussion

30 out of the 44 letters to shareholders, and 33 out of the 53 letters to stakeholders used a two-sided rhetorical strategy. Thus, at first glance, more than two thirds of the introductory letters met the audience's requirements, addressing both positive and negative aspects of the corporate activity. In total, there were 114 two-sided rhetorical moves in the letters to shareholders (cf. Table no. 1), and 77 in the letters to stakeholders (cf. Table no. 2) – a difference that may have resulted from the different length of the two types of documents, the letter to shareholders being generally longer. However, of interest for this study was the relative proportion between refutational and nonrefutational rhetorical moves in each corpus of letters, and not the total number of two-sided rhetorical moves.

Refutational two-sided moves outnumbered nonrefutational two-sided moves in both corpora, though more visibly in the letters to shareholders (82 to 32 tokens) than in the letters to stakeholders (43 to 34 tokens). As previously argued, considering the accounting exigencies of the reporting context, this proportion seemed advantageous for the persuasiveness of the corpus documents. A more precise estimation of the persuasive and the credibility impacts of these distributions on each category of introductory letters was possible after a thorough consideration of the structure and the reasonableness of the main two-sided rhetorical patterns identified in the two corpora. Here are the results of this analysis.

Letters to shareholders (from annual reports)	Non-refutational moves						Indirect refutational moves (Concessions)						Direct refutational moves (Refutation proper)						Total moves							
	External attribution		Internal attribution		No (or mixed) attribution		External attribution		Internal attribution		No (or mixed) attribution		External attribution		Internal attribution		No (or mixed) attribution									
	L	M	L	M	L	M	L	M	L	M	L	M	L	M	L	M	L	M								
Letters (L) Moves (m)																										
Past events (Retrospective sense-making)	8	11	5	5	4	7	1	2	8	14	3	3	3	5	-	-	-	-	-	-	-	-	47			
Future events	5	7	1	2	-	-	19	27	6	7	6	7	3	6	1	1	1	1	1	1	2	2	59			
Continuous issues	-	-	-	-	-	-	1	1	1	1	1	2	2	4	-	-	-	-	-	-	-	-	8			
Total moves	18		7		7		30		22		12		15		1		1		2		2		114			
	32						64						18													

Table 1. Distribution of two-sided rhetorical moves in the corpus of introductory letters to shareholders (from annual reports)



Letters to stakeholders (from CSR reports)	Non-refutational moves			Indirect refutational moves (Concessions)			Direct refutational moves (Refutation proper)			Total moves								
	External attribution	Internal attribution	No (or mixed) attribution	External attribution	Internal attribution	No (or mixed) attribution	External attribution	Internal attribution	No (or mixed) attribution									
											L	m	L	m	L	m		
<i>Letters (L)</i> <i>Moves (m)</i>	L	m	L	m	L	m	L	m	L	m								
Past events (retrospective sense-making)	2	2	6	6	12	14	-	-	5	5	1	1	-	-	28			
Future events	2	2	5	8	-	-	6	6	2	3	-	-	2	2	1	1	23	
Continuous issues	1	1	1	1	-	-	-	-	7	8	4	5	4	5	1	2	4	26
Total moves	5	15	14	14	34	28	6	16	6	7	3	5	15	77				

Table 2. Distribution of two-sided rhetorical moves in the corpus of introductory letters to stakeholders (from CSR reports)

#### 4.1 *Two-sided rhetorical moves in the letters to shareholders*

The most numerous two-sided rhetorical moves from the letters to shareholders were the concession (63 tokens), which basically followed two main rhetorical patterns. The first pattern acknowledged certain internal weaknesses of the company, and then tried to minimize their relevance in relation to the overall corporate performance. 14 concessions acknowledged past unfavourable results, then tried to alter their connotation (e.g. strategically reframing them as “in target”) or relevance (by comparison with the totality of the results, or mentioning some remediation). Similarly, the risks coming from inside the corporation were countered with remedial actions in course or – less effectively – with a commitment to remediation.

The most visible concessive (and, in general, two-sided rhetorical) pattern of this category of letters (25 moves in 18 letters) acknowledged some external risks to the future corporate performance, then, tried to offset them by summarizing the main internal strengths of the company (e.g. assets, portfolio, strategy, etc.). These moves were mainly related to the concluding stage of the managerial messages, appearing as a natural precaution in front of the unpredictability of the future (Example 1):

- (1) ...our Group is embarking on its 136th year with growing assets and a solid equity and financial structure. We are ready to take up new challenges, albeit in a macroeconomic and market context that is far from easy. (Pirelli & C. S.p.A. Directors' report 2007: 13)

The next category of two-sided rhetorical moves, nonrefutational, were mostly related to past shortcomings (23 out of 32 tokens) and, to a smaller extent (7 moves, out of which one was rhetorical), to long-standing and unavoidable (usually external) threats for company profits. These moves did not try to alter the perceived importance of the negative events, but manoeuvred strategically the perception of the corporate responsibility for their occurrence. Only in seven cases (mostly safety incidents) managers overtly assumed the company's responsibility for the negative events, expressing its compassion (in case of fatalities) and commitment to remediation, bolstered by safety standards in place and previous accomplishments. But most of the time (18 tokens), the negative outcomes were attributed to causes that were external to the company (governments, market falls, competitors, providers, even customers). Some of these factors were strategically attacked and used as scapegoats for the corporate losses,

being less likely to endanger managements' perceived capacity to control the operational environment. In Example 2, for instance, a manager reframes the role of authorities – from combatant to enabler of an illegal activity affecting the company – by strategically selecting the starting points of his argumentation (i.e., the causes of the respective illegal activity).

- (2) Illicit trade remains a key challenge and is fuelled by many factors including tax-driven price increases, poor enforcement of borders, weak laws, loosely regulated Free Trade Zones, a lack of intellectual property rights protection and the growth of the internet as a trading medium. [...] We remain fully committed to tackling the problem, and are liaising with governments, customs and law enforcement agencies around the world. (British American Tobacco (BAT) Annual Report and Accounts 2007(b): 11)

The justificatory strength of such external attributions – that eventually aim to refute an alleged managerial incapacity in meeting shareholders requirements – depends on the soundness of their supporting argumentation. Example 3, for instance, presents a questionable attempt of scapegoating, where a manager tries to blame a scientific journal and the media for the company's financial loss.

- (3) [I]n May 2007, an article in the New England Journal of Medicine (NEJM) suggested that there may be cardiovascular risk associated with Avandia, our second largest product. This was followed by intense media coverage and despite our efforts to explain the entirety of the data, which did not confirm this risk, doctors were reluctant to prescribe Avandia for new patients without further FDA guidance. [...] Sales of Avandia dropped significantly and this had a negative impact on our share price. [...] Following clarification from the FDA in October 2007, we now have a new approved label and can move ahead with more clarity. (Glaxo-SmithKline Annual Review 2007(f): 6)

The manager relies in his argumentation on the implicit assumption that risks signalled by limited data are irrelevant. However, the soundness of this assumption depends on the nature of the *parts - whole* relationship between the data used by the scientific article and “the entirety of data” provided by the company – a piece of information that is missing from the letter. For example, if the risk was erroneously inferred in the article because essential data was missing from the reasoning, and, later, the entirety of data did not confirm the risk, then the existence of a risk could be denied, and the NEJM article (and its consequences) could be considered unreasonable. But if the risk was correctly inferred (although from a limited number of cases), and the entirety of data just diminished its probability to an “acceptable” level,

then the existence of a risk could not be denied. In this case, considering the life-threatening consequences of that risk, media's and doctors' reaction might not have been unreasonable, and the company would have been responsible for the resulting financial loss. And the manager's own statements seem to go in this direction; by mentioning that the FDA has required a new label for that medicine, he implicitly admits the presence of some risk-related issues within the original label. If so, his attempt to shift the responsibility for the financial losses would not only be potentially manipulative, by ambiguity, but also inconsistent.

Finally, direct refutations – less numerous in this corpus (19 tokens) – were mostly used for defending the managerial strategy and the investment value of the company (Example 4).

- (4) Should we sell NBCU? The answer is no! I just don't see it happening... [...] It doesn't make sense. The business has outperformed its competition and GE average for the last 20 years. Content is increasing in value [...]. We are in a good cycle [...]. This is true now, and its will be true in the long term as well. (GE Annual Report 2007: 6)

The most visible rhetorical pattern of this category (6 tokens) was the dissociation between (management's) knowledge and (critics') opinion.<sup>13</sup> For example, in a sequence of proleptic argumentation belonging to the *locus from the parts and the whole* and *the locus from difference* (Example 5), the manager refutes the reasonableness of the investors' anticipated reluctance towards a profitable branch of the company, opposing the “real” value of the business to the investors' *a priori* evaluation.<sup>14</sup> He refutes the starting point of their supposed uncritical transfer of image from sector (the *whole*) to the company (the *part*) – more precisely, the implicit *endoxon* that all the companies of a sector are equally profitable.

- (5) [...] in 2007, “financial services” took on a negative connotation for investors. The equity value of banks and consumer finance companies declined [...]. Conversely, our financial services earnings grew [...]. Our financial services businesses are inherently more valuable than those of traditional banks or other services companies. Why? Because we have significant global origination in end-user

13 The “knowledge/opinion” dissociative pattern was first signalled in corporate crisis communication by Hearit (1994).

14 *Proleptic argumentation* is a rhetorical device consisting in “anticipat[ing] and answering of an objection or opposed argument before one's opponent (the audience) has put it forward” (Walton 2008: 144).

markets that we understand better than others. Because we have deep expertise [...] sound risk principles [...] great global position [etc.]. (*ibid.* 6)

Viewing these results from the perspective of their impact, it can be hypothesized that the pronounced prevalence of refutational two-sided moves in this corpus is beneficial for the communicative effectiveness of the letters to shareholders (in terms of persuasiveness and credibility). Being mostly backed by factual data, direct refutations are likely to provide a compelling defence of the managerial expertise, contributing to the overall effectiveness of the letters. Concessions, though, seem to have a particular advantage in this context (suggested also by their high frequency of occurrence), due to their ambivalent nature. By their nonrefutational component – i.e., the acknowledgment of the possible risks to shareholders' objectives – these moves may enhance management's credibility and perceived professionalism (e.g. honesty, accuracy and precaution in the evaluation of the operational environment). At the same time, by (soundly) refuting the consequences of the respective negative aspects, they may reinforce management's perceived capacity of control, thereby contributing to the persuasiveness of their account.

Consistently with previous studies in the field, the data analysis confirmed for this corpus of letters a "hedonic bias" in the managers' causal attributions. However, confirming audience's potential scepticism, external attributions are unlikely to provide a rhetorical advantage, especially when they appear in justifications within nonrefutational two-sided rhetorical moves. In these cases, the attempt to exonerate the management for the negative outcomes may conflict with its perceived capacity to control the operational environment.

Data analysis also suggests that a possible moderating role of *time* (the time-orientation of the standpoints and of the two-sided rhetorical moves) might be worth of further investigation. For example, a nonrefutational acknowledgment of a risk might weigh differently in investors' decision making than a non-refuted past negative event. Also, a future-oriented standpoint might limit the effectiveness of proper refutations, but may increase the effectiveness of concessions.

Part of these observations were also confirmed for the corpus of letters to stakeholders, but data analysis has also revealed significant differences, presented in detail in the following section.

#### 4.2 *Two-sided rhetorical moves in the letters to stakeholders*

In the letters to stakeholders (in CSR reports), the distribution of two-sided rhetorical moves followed a different pattern (cf. Table no. 2). A first difference regarded time-orientation, remarkable being the extensive coverage of issues affecting the companies on a continuous basis – 24 out of 77 tokens, compared to 8 out of 114 tokens in the letters to shareholders. A second difference was the more balanced distribution of the three types of two-sided rhetorical moves, and the prevalence of nonrefutational moves (34 tokens) over concessions (28 tokens) and direct refutations (15 tokens). However, 13 similar nonrefutational two-sided moves were found in letters that belonged to subsidiaries of the same company; so, this could be a company- or industry-specific pattern that might have biased the primacy of nonrefutational moves in the corpus. A third difference regarded managers' attributional style in relation to the acknowledged shortcomings (in both nonrefutational and concessive two-sided moves). Contrary to the letters to shareholders, the letters to stakeholders show prevalence of internal attributions (31 internal attributions compared to 11 external attributions) – a possible benefit for the credibility of these letters. Illuminating for a better understanding of the potential persuasive and credibility effects of this distribution was the analysis of the main rhetorical pattern encountered in the corpus.

The most visible rhetorical pattern of the letters to stakeholders was the nonrefutational acknowledgment of past safety incidents (19 tokens). Only in six cases did managers admit company responsibility for the incidents. In the other 13 cases, they made no reference to the causes of the incidents – possibly, due to legal considerations, given the recency of the events – limiting themselves to statements of regret (in case of life-losses) and a commitment to remediation. A common strategic manoeuvre within this rhetorical pattern – consistent with previous observations on the “Pollyanna Hypothesis” in the letters to shareholders (Thomas 1997)<sup>15</sup> – was to positively reframe safety shortcomings by placing emphasis on their overall incidence rate (usually in progress compared to the previous year) rather than on single incidents (Example 6).

15 The ‘Pollyanna Hypothesis’ refers to managers’ propensity to employ positive language regardless of the financial state of the company. (Hildebrandt & Snyder 1981, *in* Thomas 1997: 48–49)

- (6) In pursuit of our zero harm objective, [our company] achieved a 16,2% reduction in the Total Recordable Injury Frequency Rate (Xstrata Copper Ernest Henry Mining Sustainability Report 2007: 2)

A second category of two-sided nonrefutational moves (eight tokens) anticipated various challenges for the companies, without mentioning whether or how these challenges would be overcome. When related to job cuts, these moves included also a justification and mentioned various reparatory actions, bolstered with claims of responsible corporate behaviour towards the employees (Example 7):

- (7) The company restructuring programme announced in 2007 will help us remain a competitive and sustainable business. These changes are necessary but have inevitably required us to reduce employee numbers. We aim to treat our employees with dignity and respect and offer a wide range of support for all affected staff. (GlaxoSmithKline Corporate Responsibility Report 2007 (b): 7)

Arguably the most critical component of this rhetorical pattern, the *justification* should prove – in line with Schütz (1998: 619) – that the negative event was inevitable, being necessary for the achievement of a more important corporate objective; therefore, the company should not be blamed. The justification presented in Example 7 relies on a common domain-specific assumption (*endoxon*) that the reduction of personnel helps companies to maintain their competitiveness and business sustainability. However, since the reduction of personnel is only one way of reaching these objectives, this *endoxon* might not be unanimously shared by the stakeholders (e.g. by the employees); therefore, this justification might be seen as neither credible nor persuasive.

In what regards the refutational two-sided rhetorical moves (43 tokens), they were divided between concessions (28 tokens) and direct refutations (15 tokens).

Two concessive patterns were similar to patterns observed in the letters to shareholders, being employed in relation to past sustainability shortcomings (5 moves) and to (usually external) risks for the future sustainability of the company (6 moves). But the distinctive feature of the two-sided refutational moves from the letters to stakeholders was the extensive coverage of the issues affecting the companies on a continuous basis – 24 from 43 tokens, compared to 8 from 82 in the letters to shareholders.

Particularly relevant for the credibility of the letters to stakeholders were the two-sided refutational moves that defended the position of the companies on *matters of principle* related to their voluntary agency towards the stakeholders. Some of these moves reinforced the agency of the company; while others set limits to it. Thus, two concessive moves defended the very possibility of a company to be sustainable, in spite of the counter-attitudinal connotations of its industrial profile – e.g. oil and gas (BP), tobacco (BAT). Other concessive moves (4 tokens) defended the actual sustainability of the companies with “risky” profiles (e.g. fuel, energy production), relying on factual arguments (like on-going action plans) and/or on value-based arguments (e.g. commitments, corporate values and business principles). Sometimes, the commitment to principles itself was defended (Example 8):

- (8) [We] have established principles for the climate change debate consistent with [...] our company’s values. [For example] we are a large independent power generator, but our principles [...] do not promote free emissions allowances under a cap-and-trade program to power generators. [...] We recognize that it could be argued that our principles are flawed. The company would be better off supporting free allowances for all generators based on output. But that’s why they call them principles. (Entergy Sustainability Report 2007: 3)

Value-based arguments rely on an ontological principle (*maxim*) that states that the quality of the cause may influence the quality of the effect. In this example, for instance, the sustainability entailed by the corporate values (the *final cause* of the corporate behaviour) that underlie those business principles (as *formal cause* of the corporate behaviour) may bring about a sustainable corporate behaviour. However, considering the factual exigencies of the corporate reporting, such arguments might constitute an insufficient defence and be potentially misleading – unless supported by relevant facts. That is because, in reality, they entail a commitment to how things *should be* (a deontic argument), and not a course of action, i.e., how things *actually are* (an ontological argument).<sup>16</sup>

Representative for the letters to stakeholders were also the refutational patterns that addressed a major CSR topic – the historically assumed opposition between shareholders’ and stakeholders’ interests, and the debate about their priority for the corporate management (Smith 2003). There were

16 More on this distinction and on the use of modalities in argumentation can be found in Rocci (2008). An example of the potentially manipulative exploitation of the two modalities in CSR reporting is presented in Filimon (2009).



two main (contrasting) approaches to this issue. The first approach – basically concessive – aimed to prove that the two categories of interests are not mutually exclusive, companies being able to operate socially responsibly in spite of the acknowledged priority of shareholders' interests. One argument in this direction was the moral sense of the employees (the *efficient cause* of the corporate behaviour), who “have their own good sense about the right thing to do [and would not] want to work for a company that only cares for profits” (*BAT Sustainability Report 2007*: 3). Thus, the social responsibility of the company would result from the social responsibility of its employees – a rationale entailed by the *locus from the whole and the parts* and the *locus from the efficient cause*. A second argument in this direction – met in two concessive moves (Example 9) and one refutational move based on dissociation (Example 10) – was the potential contribution of corporate social responsibility to the achievement of the companies' financial goals:

- (9) Our [company's] goal is to create sustainable shareholder value. [...] But if a business wants to succeed long term, it makes absolute sense for it to operate responsibly. (*BAT Sustainability Report 2007*: 3)
- (10) [Reporter]: So sustainability contributes to growth and value creation.  
[President]: Absolutely. Initially people thought of it as a cost factor, which indeed it is when you treat it as an add-on. However, if it's designed into the way you do things from the beginning as it is here at Philips, it saves you money because you're operating more effectively. So today we recognize that sustainability offers significant business opportunities. (*Philips Sustainability Report 2007*: 8)

This “win-win” rhetorical pattern relies on a *maxim* of the *locus from the efficient cause* and the *locus from the final cause* that states that “an agent's commitment is reliable if it is bound to its strongest interest” (Rigotti, Greco Morasso, Palmieri & Palmieri 2007, in Filimon 2011: 471). However, being expressed in a deontic modality, the pro-attitudinal argument of Example 9 is less effective in supporting the actual responsibility of the company than the ontological pro-attitudinal argumentation of Example 10.

The second approach to the “stakeholders-shareholders” debate, met in five moves, was essentially refutational. Managers acknowledged the priority of the shareholders' financial interests in order to set limits to the stakeholders' expectations towards the company (Example 11) and refute the opinion – “shared by many” – that corporations should be “all things to all stakeholders” (*GE Citizenship Report 2007–2008*: 11).

- (11) Businesses have to be honest about what they are and what they can do. Our goal is to create sustainable shareholder value. Businesses can't assume the role of governments, charities, political parties, action groups or the many other bodies that make up society. (BAT Sustainability Report 2007: 3)

Example 11 relies on a *maxim* of the *locus from the final cause*, according to which an agent cannot be reasonably expected to assume roles that are not related to its final goals (Filimon 2011: 472). Other dissociative attempts were less reasonable. In Example 12, a manager fallaciously assumes that in a multi-agent system, the smaller the size of an agent, the smaller is its influence on the effect produced by the system:

- (12) With healthcare, there are multiple issues and more stakeholders. Our contribution overall is modest when you consider that we are just a relatively small part of the total healthcare value chain, in addition to doctors, hospitals, pharmaceutical companies, insurers, etc. (Philips Sustainability Report 2007: 10)

By strategically dissociating and (re)defining in a self-serving manner the roles in social responsibility, managers were able to claim the fulfilment of the corporate social duty, and to eventually attack the other social co-agents for not having fulfilled *their* part of responsibility – that is, the role they were strategically assigned by the managers themselves (Example 13):

- (13) As a business we need to ensure that everything we do reflects what our customers want and need. We must make healthcare accessible and affordable for everyone, not only in advanced markets but also in new and emerging markets. Governments and insurers must also do their part. [...] I believe that a marriage of stakeholder needs and meaningful, cohesive policies will lie at the heart of successful strategies to meet the healthcare challenge. (*ibid.*)

The pragmatic rationale of Example 13 relies on the same argumentative *locus* as Example 12, but on a different *maxim*, according to which if a necessary effect requires the collaboration of multiple agents, then all agents must do their part in order to produce that effect. There are, however, a number of weak points in this attack. Although the necessity to make healthcare accessible for everyone is a common-sense *endoxon*, the acceptance of the other starting points of the manager cannot be taken for granted – e.g. his unilateral opinions about who should do what in making healthcare accessible. But in the wider context of CSR reporting, this attack could be considered, in essence, a strategic manoeuvre based on *concordatio*. Governments and insurers are important stakeholders of the company, and the collaboration between corporate and non-corporate agents is

recommended by stakeholders themselves – that is., by the initiatives that represent stakeholders' interests (e.g. UNCG 2010, GRI 2004–2011). Thus, the necessity of cooperation, *in se*, can be considered an effective starting point for this attack – although the same does not hold for the manner of the cooperation.

As these examples reveal, the marked credibility issue of the CSR reports (detailed in section 1) is plainly reflected in this corpus of letters by the large space allotted to the defence of the corporate positions in agency-related matters of principle, and by the complexity of the corresponding argumentation. A persuasive advantage of the extensive use of two-sided refutational moves in relation to these topics cannot be taken for granted. The argumentative reconstruction highlighted a number of weak points and fallacies, especially in direct refutations (e.g. self-serving definitions, an abusive use of starting points, logically unsound or insufficient arguments) that may damage the communicative effectiveness of the corporate messages (in terms of persuasiveness and credibility). However, considering the likely scepticism of the audience towards the corporate candour in relation to CSR principled topics, concessions might enjoy a credibility advantage, due to their non-refutational component. Also, due to a generally better quality of their pro-attitudinal argumentation (as compared to direct refutations), concessions might actually provide a persuasive advantage.

The large number of nonrefutational two-sided moves (many of them associated with internal attributions) may increase the credibility of the informative account of the letters to stakeholders. However, this favourable effect is likely to be offset by the potentially detrimental effect of these moves on the persuasiveness of the letters. That is because non-refutational moves mainly appeared in the “material performance” area (32 out of 34 tokens), where they significantly outnumbered the two-sided refutational moves (11 tokens), conveying thus a negative impression on the managerial performance in the field.

## 5. Conclusions

This study investigated the use of two-sided rhetorical strategies in two corpora of introductory letters to shareholders, respectively stakeholders, from annual and CSR reports.

By means of a mixed methodological approach that combined the analytical tools of argumentation theory with insights coming from the persuasion research on message-sidedness effects, the study explored the extent of use, the argumentative structures, and, *ex hypothesi*, the potential impact of two-sided rhetorical strategies on the communicative effectiveness (in terms of persuasiveness and credibility) of each category of introductory letters. This impact was estimated at corpus level, based on the relative proportion of refutational and nonrefutational two-sided rhetorical moves (identified by argumentative reconstruction) in the two corpora of letters, and on the potential credibility and persuasive effects of the two-sided rhetorical moves in the communicative context of each category of introductory letters. This estimation was then confronted to the argumentative analysis of the main two-sided rhetorical patterns identified in the two corpora of letters.

The rhetorical profiles delineated by the distribution of two-sided rhetorical moves in the two corpora of introductory letters served as a guide for seizing rhetorical tendencies, and not as a statistical indicator – that could be a subject for further research.

Data analysis showed that more than two thirds of the letters of each corpus employed a two-sided rhetorical strategy – apparently meeting the genre-specific rhetorical exigencies for an account on both positive and negative aspects of the corporate activity. The characteristics of the rhetorical situation of each category of introductory letters were reflected in the main topics addressed by the two-sided rhetorical moves in the respective corpora, and in the way managers handled the counterarguments. Performance-related topics prevailed in the letters to shareholders, while impression management topics were more visible in the letters to stakeholders, which allotted more space to the defence of the corporate *ethos*. Although a self-serving attributional bias was observed in the letters to shareholders, the prevalence and the good argumentative quality of refutational two-sided moves, especially of concessions – whose ambivalent nature might provide a

rhetorical advantage in this context – seemed beneficial in terms of persuasiveness and credibility. In the letters to stakeholders, the refutational two-sided moves prevailed as well, but mostly in relation to impression management topics, which are more likely to elicit audience scepticism towards the managers' impartiality. In this context, concessions might have a credibility advantage and, due to the better quality of their underlying argumentation, be more persuasive than direct refutations. But the prevalence of non-refutational moves in performance-related arguments seemed detrimental to the overall communicative effectiveness of these letters, a possible positive effect in terms of credibility being likely to be offset by the negative consequences on the persuasiveness of the corporate claims of responsible behaviour.

Data analysis has also suggested that a possible moderating role of *time* (the time-orientation of both standpoints and two-sided rhetorical moves) might be worth of further investigation.

This study tried to make a step towards a more comprehensive methodological approach to the evaluation of the two-sided communication, able to integrate the normative analytical rigour of argumentation theory with the explanatory and predictive strengths of the empirical research on persuasion. At the same time, by its applicative side, this investigation aimed to contribute to the study of corporate apologia and, in general, of the effectiveness of the communication between corporations and their multiple stakeholders.

## References

- Adams, C.A. (2004): "The ethical, social and environmental reporting-portrayal gap", *Accounting, Auditing & Accountability Journal* 17 (5), 731-757.
- Aerts, W. (1994): "On the use of accounting logic as an explanatory category in narrative accounting disclosures", *Accounting, Organizations and Society* 19 (4/5), 337-353.
- Bettman, J.R. & Weitz, B.A. (1983). "Attributions in the Board Room: Causal reasoning in corporate annual reports", *Administrative Science Quarterly* 28 (2), 165-183.
- Bitzer, L.F. (1992) [1968]: "The rhetorical situation", *Philosophy and Rhetoric*, Supplementary Issue, 1-14.
- Clarke, G. & Murray, L.W. (B.) (2000): "Investor relations: perceptions of the annual statement", *Corporate Communication: An International Journal* 5 (3), 144-151.

- Clatworthy, M.A. & Jones, M.J. (2006): "Differential patterns of textual characteristics and company performance in the chairman's statement", *Accounting, Auditing & Accountability Journal* 19 (4), 493-511.
- Eemeren, F.H. van (2010): *Strategic Maneuvering in Argumentative Discourse: Extending the pragma-dialectical theory of argumentation*. John Benjamins, Amsterdam / Philadelphia.
- Eemeren, F.H. van & Grootendorst, R. (2004): *A systematic theory of argumentation: The pragma-dialectical approach*. Cambridge University Press, Cambridge.
- The European Parliament and the Council of the European Union. "Directive 2006/46/EC of the European Parliament and of the Council", *Official Journal of the European Union*, L244, 1-7, eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2006:224:0001:0007:EN:PDF, last accessed August 31, 2013.
- Filimon, I.A. (2009): "Kyosei – an example of cultural keyword argumentatively exploited in corporate reporting discourse", *Studies in Communication Sciences* 9 (2), 131-151.
- (2011): "Argumentative valences of the key-phrase *value creation* in corporate reporting", in Eemeren, F.H. van, Garssen, B.J., Godden, D. & Mitchell, G. (eds), *Proceedings of the 7th Conference of the International Society for the Study of Argumentation* (CD-ROM). Sic Sat, Amsterdam / Rozenberg, 461-479.
- Financial Reporting Council (FRC) (2011): "True and Fair". Online paper, www.frc.org.uk/FRC-Documents/FRC/Paper-True-and-Fair.aspx, last accessed August 31, 2013.
- Global Reporting Initiative (GRI) (2000-2011): "RG. Sustainability Reporting Guidelines", Version 3.1., www.globalreporting.org/resourcelibrary/G3.1-Sustainability-Reporting-Guidelines.pdf, last accessed August 31, 2013.
- Hearit, K.M. (1994): "Apologies and public relations crises at Chrysler, Toshiba, and Volvo", *Public Relations Review* 20 (2), 113-125.
- (2006): *Crisis Management by Apology: Corporate responses to allegations of wrongdoing*. Lawrence Erlbaum Associates, Mahwah, NJ / London.
- Hildebrandt, H. H. & Snyder, R. (1981): "The Pollyanna hypothesis in business writing; initial results, suggestions for research", *The Journal of Business Communication* 18 (1), 5-15.
- Hulle, K. van (1993): "Truth and untruth about true and fair: A commentary on 'A European true and fair view' comment", *European Accounting Review* 1, 99-104.
- Hyland, K. (1998): "Exploring corporate rhetoric: Metadiscourse in the CEO's letter", *The Journal of Business Communication* 35 (2), 224-245.
- Kirk, N. (2006): "Perceptions of the *true and fair view* concept: An empirical investigation", *Abacus* 42 (2), 205-235.
- Merkel-Davies, D.M., Brennan, N.M. & McLeay, S.J. (2011): "Impression management and retrospective sense-making in corporate narratives: A social psychology perspective", *Accounting, Auditing & Accountability Journal* 24 (3), 315-344.
- O'Keefe, D.J. (1999): "How to handle opposing arguments in persuasive messages: A meta-analytic review of the effects of one-sided and two-sided messages", in Roloff, M.E. (ed.) & Paulson G.D. (ed. assist.), *Communication Yearbook*. 22, SAGE Publications Oaks / London / New Delhi, 209-249.
- Petty, R.E. & Cacioppo, J.T. (1986): *Communication and persuasion: Central and peripheral routes to attitude change*. Springer, New York / Berlin / Heidelberg / London / Paris / Tokyo.
- Rigotti E. (2006): "Relevance of context-bound *loci* to topical potential in the argumentation stage", *Argumentation* 20, 519-540.

- Rigotti, E. & Greco Morasso, S. (2010): "Comparing the Argumentum Model of Topics to other contemporary approaches to argument schemes: The procedural and material components", *Argumentation* 24 (4), 489-512.
- Rigotti, E. & Rocci, A. (2006): "Towards a definition of the communication context: Foundations of an interdisciplinary approach to communication", *Studies in Communication Sciences* 6 (2), 155-180.
- Rocci, A. (2008): "Modality and its conversational backgrounds in the reconstruction of argumentation", *Argumentation* 22, 165-189.
- Rogers, J.L. & Stocken, P.C. (2005): "Credibility of Management Forecasts", *The Accounting Review* 80 (4), 1233-1260.
- Rowland, R.C. & Jerome, A.M. (2004): "An organizational apologia: A reconceptualization", *Communication Theory* 14 (3), 191-211.
- Rutherford, B.A. (2005): "Genre analysis of corporate annual report narratives: A corpus-linguistics-based approach", *Journal of Business Communication* 42 (4), 349-378.
- Salancik, G.R. & Meindl, J.R. (1984): "Corporate attributions as strategic illusions of management control", *Administrative Science Quarterly* 29, 238-254.
- Schaltegger, S. & Burritt, R. (2006): "Corporate sustainability accounting: A catchphrase for compliant corporations or a business decision support for sustainability leaders?", in Schaltegger, S., Bennet, M. & Burritt, R. (eds), *Sustainability Accounting and Reporting*. Springer, Dordrecht, 37-59.
- Schütz, A. (1998): "Assertive, offensive, protective, and defensive styles of self-presentation: A taxonomy", *The Journal of Psychology* 132 (6), 611-628.
- Sampson, C. (1985): "Fair Presentation". Address to American Accounting Association, Northeast Regional Meeting. Syracuse University, Syracuse, New York, [www.sec.gov/news/speech/1985/041985sampson.pdf](http://www.sec.gov/news/speech/1985/041985sampson.pdf), last accessed August 31, 2013.
- Schoonraad, N., Grobler, A.F. & Gouws, D.G. (2005): "An inclusive and integrated approach to financial communication: A conceptual model", *Communication: South African Journal for Communication Theory and Research* 31 (2), 273-294.
- Skulstad, A.S. (2005): "The use of meta-discourse in introductory sections: if a new genre", *International Journal of Applied Linguistics* 15 (1), 71-86.
- Smith, H. J. (2003): "The shareholders vs. stakeholders debate", *MIT Sloan Management Review* 44 (4), 85-90.
- Snoeck Henkemans, A.F. (1997): *Analyzing Complex Argumentation: The reconstruction of multiple and coordinatively compound argumentation in a critical discussion*. Sic Sat, Amsterdam.
- (1995): "'But' as an indicator of counter-arguments and concessions", *Leuense Bijdragen* 84, 281-293.
- Stanton, P. & Stanton, J. (2002): "Corporate annual reports: Research perspectives used", *Accounting, Auditing & Accountability Journal* 15 (4), 478-500.
- Stapleton, K. & Hargie, O. (2011): "Double-bind accountability dilemmas: Impression management and accountability strategies used by senior banking executives", *Journal of Language and Social Psychology* 30 (3), 266-289.
- Thomas, J. (1997): "Discourse in the marketplace: The making of meaning in annual reports", *The Journal of Business Communication* 34 (1), 47-66.
- Tsang, E.W.K. (2002): "Self-serving attributions in corporate annual reports: A replicated study", *Journal of Management Studies* 39 (1), 51-65.

- United Nations Global Compact (UNGC) (2010): “Blueprint for Corporate Sustainability Leadership”, [www.unglobalcompact.org/docs/issues\\_doc/lead/Blueprint\\_english.pdf](http://www.unglobalcompact.org/docs/issues_doc/lead/Blueprint_english.pdf), last accessed August 31, 2013.
- Walton, D. (2006): *Fundamentals of Critical Argumentation*. Cambridge University Press, New York.
- (2008): “Proleptic argumentation”, *Argumentation and Advocacy* 44, 143-154.
- Winsor, D. (1993): “Owning corporate texts”, *Journal of Business and Technical Communication* 7 (2), 179-195.

## APPENDIX 1.

### Corpus of introductory letters from annual reports

- Akzo Nobel N.V. *Annual Report 2007*. Chairman’s statement.
- Audi. *2007 Annual Report*. Editorial.
- BNP Paribas. *Annual Report 07*. Chairman’s and Chief Executive Officer’s statement.
- British American Tobacco (BAT). *Annual Report and Accounts 2007*. (a) Chairman’s statement. (b) Chief Executive’s Review.
- British Petroleum (BP) plc. *BP Annual Report and Accounts 2007*. (a) Chairman’s Letter. (b) Group chief executive’s review.
- Canon Inc. *Canon Annual Report 2007*. (a) To our stockholders. (b) Message from the President.
- Entergy Corporation. *2007 Annual Report*. To our stakeholders.
- General Electric Company. *GE Annual Report 2007*. Letter to investors.
- GlaxoSmithKline plc. *Annual Review 2007*. (a) Chairman and CEO Summary. (b) Q&A: Strategy. (c) A personal perspective from Mike Thomson. (d) Q&A: Consumer healthcare. (e) A personal perspective from Ken James. (f) Q&A: Company performance. (g) A personal perspective from Duncan Learmouth. (h) Q&A: Research and Development. (i) A personal perspective from Dr. Jingwu Zang. (j) Q&A: Community partnerships. (k) A personal perspective from Justine Frain.
- Heineken N.V. *2007 Annual Report*. Chief Executive’s Statement.
- KBC Group N.V. *Annual Report 2007*. Statement by the Chairman of the Board of Directors and the President of the Executive Committee.
- Melco International Development Limited. *Annual Report 2007*. Chairman & CEO’s statement.
- Nestlé S.A. *Management Report 2007*. Letter to our shareholders.
- PepsiCo, Inc. *2007 Annual Report*: (a) Letter to Shareholders. (b) A perspective from our Chairman and CEO.
- Pirelli & C. S.p.A. *Directors’ report and consolidated financial statement at December 31, 2007*. Chairman’s Letter.
- Pirelli & C. Real Estate S.p.A. *Annual Report 2007*. Message from the CEO.
- The Procter & Gamble Company (P&G). *2007 Annual Report*. Letter to shareholders.
- Royal Philips Electronics. *Philips Annual Report 2007*. Message from the President.
- SkodaAuto. *2007 Annual Report*. Letter from the Chairman.
- Sinopec Corp.–China Petroleum & Chemical Corporation. *2007 Annual Report and Accounts*. Chairman’s Statement.
- Vodafone Group plc. *Annual Report for the year ended 31 March 2008*. (a) Chairman’s Statement. (b) Chief Executive’s Review.
- Volkswagen AG. *Annual Report 2007*. (a) Letter to our Shareholders. (b) Talking to.
- Volkswagen Bank GmbH. *Annual Report 2007*. Foreword of the Board of Management.
- Volkswagen Financial Services AG. *Annual Report 2007*. Foreword of the Board of Management.
- Volkswagen Leasing GmbH. *Annual Report 2007*. Foreword of the Board of Management.
- Xerox Corporation. *Annual Report 2007*. Letter to shareholders.
- Xstrata plc. *Annual Report 2007*. (a) Chairman’s Statement. (b). Chief Executive’s Report.



## APPENDIX 2.

## Corpus of introductory letters from CSR reports

- Akzo Nobel N.V. *Sustainability Report 2007*. Chairman's statement.
- BNP Paribas. *Report on Environmental and Social Responsibility 07*. Statement from the Chairman and Chief Executive Officer.
- British American Tobacco (BAT). *Sustainability Report 2007*. Chief Executive's overview.
- British Petroleum (BP) plc. *Sustainability Report 2007*. Group chief executive's introduction.
- BP Australia Pty Limited. *Sustainability Review 2007-08*. BP Australia president's introduction.
- BP Azerbaijan SPU. *Sustainability Report 2007*. Foreword by the president of BP Azerbaijan Strategic Performance Unit (SPU).
- BP Georgia. *Sustainability Report 2007*. A view from our country manager.
- Canon Inc. *Sustainability Report 2008*. Message from Top Management.
- Entergy Corporation. *Sustainability Report 2007*. To our stakeholders.
- General Electric Company. *GE Citizenship Report 2007-2008*. (a) A letter from Jeff Immelt. (b) A letter from Brackett Denniston and Bob Corcoran.
- General Electric Company. *GE ecomagination Report 2007*. To our investors, customers and other stakeholders. (A letter from Jeff Immelt and Lorraine Bolsinger).
- GlaxoSmithKline plc. *Corporate Responsibility Report 2007*. (a) Q&A with Chairman. (b) Letter from the CEO.
- Heineken N.V. *2007 Sustainability Report*. Foreword.
- KBC Group N.V. *Corporate Social Responsibility Report 2007*. Word of welcome.
- Melco International Development Limited. *Corporate Social Responsibility Report 2007*. Foreword: Message from the Group Chairman & CEO.
- Nestlé S.A. *The Nestlé 'Creating Shared Value' Report 2007*. Creating Shared Value: the role of business in society.
- PepsiCo, Inc. *Overview 2007*. Sustainability Q&A with Indra Nooyi, PepsiCo Chairman & CEO.
- Royal Philips Electronics. *Sustainability Report 2007*. Interview with the President.
- Pirelli & C. S.p.A. *Sustainability Report*. In: *Pirelli & C. S.p.A., Directors' report and consolidated financial statement at December 31, 2007*, 177-296. [Introductory message from the Chairman]
- Pirelli & C. Real Estate S.p.A. *Sustainability Report 2007*. Message from the CEO to stakeholders.
- The Procter & Gamble Company (P&G). *2007 Global Sustainability Report*. CEO Statement.
- SkodaAuto. *Sustainability Report 2007/2008*. Fore-word of the Chairman of the Board of Directors.
- Sinopec Corp.-China Petroleum & Chemical Corporation. *Sustainable Development Report 2007*. Welcome from Chairman.
- Vodafone Australia Limited. *Corporate Responsibility Report 2008*. Letter from Russ.
- Vodafone Greece. *Corporate Responsibility Report April 2007- March 2008*. Letter from the Chairman of the Board of Directors and Chief Executive.
- Vodafone Group Plc. *Corporate Responsibility Report for the year ended 31 March 2008*. Message from the Chief Executive.
- Vodafone Malta Limited. *Corporate Social Responsibility April 2006- March 2008*. Letter from the Chief Executive.
- Vodafone New Zealand Limited. *Corporate Responsibility Report 2008 for the 2008 Financial Year. Open: 10 Years of Vodafone New Zealand*. CEO Introduction.
- Vodafone Portugal - Comunicações Pessoais S.A. *Social Responsibility Report 07*. Chairman's message.
- Vodafone Spain. *Corporate Responsibility Report 2007-08: Responsible Innovation*. Letter from the Chairman CEO.
- Vodafone UK. *CR Review. Vodafone UK Corporate Responsibility 2007/08*. CEO Q&A.
- Volkswagen AG. *Sustainability Report 2007/2008: We are moving into the future responsibly*. (a) Chairman's Foreword. (b) Foreword from the Works Council.
- Xerox Corporation. *2007 Report on Global Citizenship: Our Word. Our Work. Our World*. Chairman's letter.
- Xstrata plc. *Sustainability Report 2007*. Chief Executive's Report.
- Xstrata Copper-Canada. *Sustainability Report 2007*. Chief Operating Officer's Report.
- Xstrata Copper-El Pachón Project. *Sustainability Report 2007*. (a) Message from the Executive General Manager Project Development Division. (b) Message from the General Manager of El Pachón.
- Xstrata Copper-Ernest Henry Mining. *Sustainability Report 2007*. General Manager's message.
- Xstrata Copper-Frieda River Project. *Sustainability Report 2007*. General Manager's message.

- Xstrata Cooper - Minera Alumbrera. *Sustainability Report 2007*. (a) Chief Executive's Message. (b) Message from the General Manager.
- Xstrata Copper-North Chile Division. *2007 Sustainability Report*. Message from the Chief Operating officer.
- Xstrata Copper-Southern Peru Division-Las Bambas Mining Project. *Sustainability Report 2007*. Report from the Xstrata Copper Southern Peru Division's Chief Operating Officer.
- Xstrata Copper-Southern Peru Division-Xstrata Tintaya. *Sustainability Report 2007*. Report from the Xstrata Copper Southern Peru Division's Chief Operating Officer.
- Xstrata Copper-Townsville. *Sustainability Report 2007*. General Manager's Message.
- Xstrata Mount Isa Mines. *Sustainability Report 2007*. Executives' Message.
- Xstrata Nickel. *Making Sustainable Progress*. Our Position.
- Xstrata South Africa (PTY) Limited-Alloys Division. *Sustainability Report 2007*. CEO's Review.
- Xstrata Zinc North Queensland. *Sustainability Report 2007*. (a) Chief Executive's Message. (b) Executive General Managers Message.

# Presupposing redefinitions

Fabrizio MACAGNO, Universidade Nova de Lisboa

## 1. Introduction – Definitions in argumentation

Words, and in particular “ethical” or “emotive” terms (Stevenson 1937: 18-19), are extremely powerful instruments. They can be used to modify our beliefs, our knowledge and our point of view, but also to conceal states of affairs (Schiappa 2003) to influence our judgments and decisions. More importantly, by changing the meaning of a word it is possible to modify the way reality is perceived by our interlocutor.

Zarefsky (1998) and Schiappa (2003: 111-112; 130) pointed out the implicit dimension of this act of naming reality, which they call “argument by definition”. Instead of putting forward a classification and support it by a definitional reason, the speaker simply names reality, leaving the definition unexpressed. Instead of stating or advancing a definition, he takes it for granted, considering it as part of the interlocutors’ common ground. This move is not a simple definitional act. The speaker is not defining, in the sense that he is not proposing or stipulating a definition. He is actually doing much more. He is presupposing a redefinition, or rather he is taking for granted a new, unshared meaning.

The purpose of this paper is to show how the structure of the act of presupposing can help understand the force and the dangers of implicit redefinitions, providing an instrument to assess their reasonableness. It will be shown how the persuasiveness of certain words can be represented by combined patterns of reasoning, which depend on definitions and values. By changing the definition of a word it is possible to alter the consequent evaluative reasoning. Redefinitions are not *per se* fallacious moves; on the contrary, there are often needed, especially if the boundaries of its *definiendum* are blurred and indistinct (Sager 2000: 216-217, Walton 2005: 169-173, Gallie 1956, Sorensen 1991). However, the speaker can define or redefine by performing different types of speech acts subject to specific conditions, of which the most troublesome is the implicit act of presupposing a definition.

By describing the conditions of this *non*-act, it is possible to understand the relationship between taking a proposition for granted and presuming its acceptance. On this perspective, presumptive reasoning becomes a crucial dimension of presupposition, allowing one to assess the reasonableness of tacit moves such as the implicit redefinition.

## 2. The persuasive dimensions of words

The power of definitions and redefinitions consists in the conclusion that the *definiendum* triggers or is used to support. For instance, concepts such as *war* or *terrorism* are usually judged negatively, and can be used to arouse negative emotions or elicit negative judgments. For this reason, naming can be considered a form of condensed argument composed of two dimensions, a classification of reality and a value judgment.

The distinction between the two aspects of the persuasive force of words was drawn by Stevenson in his analysis of ethical words. On his view, ethical or emotive words were described as words having the power of directing attitudes. Stevenson noted that words such as “peace”, “war” or “democracy” are not simply used to describe and therefore affect the cognitive reaction of the interlocutor. They can provoke a different type of reaction, emotive in nature. Stevenson called these two reactions “descriptive meaning” and “emotive meaning”, defining “meaning” as a stable correlation between the sign, a stimulus, and a psychological response of the addressee (Stevenson 1944: 54). Ethical words have the power of directing attitudes, arousing emotions and suggesting, or rather recommending, courses of actions. (Stevenson 1937: 18-19): “Instead of merely describing people's interests, they change or intensify them. They *recommend* an interest in an object, rather than state that the interest already exists”.

These words have the tendency to encourage future actions (Stevenson 1938b: 49-50), and lead the hearer towards a decision by affecting his system of interests (Stevenson 1944: 210). Descriptive and emotive meanings can be interrelated or independent to each other. Some terms (for instance “peace” or “war”) have a positive or negative emotional meaning because their descriptive one refers to a state of affairs usually assessed positively or

negatively by the community of speakers (Stevenson 1944: 72). The two meanings are independent from each other in other words. For instance, the difference between “cur” and “dog”, “elderly maiden” and “old spinster” simply consists in the different emotive reaction that they arouse (Stevenson 1937: 23, Stevenson 1938a: 334-335). The power of ethical or emotive words was underscored by Stevenson, who pointed out the strict relation between definition and persuasion (Stevenson 1944: 210), claiming that “to choose a definition is to plead a cause, so long as the word defined is strongly emotive”.

Stevenson noticed that the emotive meaning of a word cannot be defined, but it can be modified by two powerful tactics, quasi-definitions and persuasive definitions. In the first case, the descriptive meaning is maintained, while the emotive one is altered. The quasi-definition does not describe the *definiendum*, simply qualifies it in order to arouse contrary or different emotions. For instance we can consider the Don Juan’s definition of “fidelity” as “being trapped forever in the same relationship and as good as dead from youth onwards to the other pretty faces that might catch our eye!” (Molière 2000: 98) or the following definition of “peace” (Bierce 2000: 179):

1. Peace. In international affairs, a period of cheating between two periods of fighting

In these cases, the speaker describes what the words is commonly used to refer to, and qualifies it using epithets or metaphors eliciting negative instead of positive evaluations.

Persuasive definitions are much more powerful and dangerous tactics. They consist in modifying the extension of a term, so that it can be used to refer to a different fragment of reality, maintaining its emotive meaning unaltered. For instance, we can consider the following redefinition of “peace”, or rather, “true peace” (Barack Obama, *Nobel Peace Prize Acceptance Address Oslo*, Norway December 10, 2009):

2. *Implicit redefinition: “Peace”*

Peace is not merely the absence of visible conflict. Only a just peace based on the inherent rights and dignity of every individual can truly be lasting. [...]A just peace includes not only civil and political rights -- it must encompass economic security and opportunity. For true peace is not just freedom from fear, but freedom from want.

Here the emotive meaning of “peace” is maintained, but its descriptive meaning is modified to include war operations. A last tactic can be considered as the combination of the two methods described by Stevenson. A word is simply renamed, so that its emotive meaning is altered by exploiting the descriptive meaning of the new *signifiant*. For instance, in order to avoid the negative value judgments triggered by a “war on terror”, under the Obama administration such military operations were simply renamed.<sup>1</sup>

3. This Administration prefers to avoid using the term “Long War” or “Global War on Terror” [GWOT]. Please use “Overseas Contingency Operation.”

The descriptive meaning of an “operation” is different from the one of “war”. As a consequence, the two concepts trigger distinct value judgments. The same state of affairs was named differently to take advantage of the “emotive” meaning associated with the new “descriptive” one.

### 3. Arguments in words

Stevenson’s account of emotive and descriptive meaning can be analyzed from an argumentative perspective as a twofold dimension of reasoning. On this view, Stevenson’s meaning, corresponding to the propensity of a word to elicit certain attitudes, can be thought of as a process of reasoning triggered or presupposed by the use of the word. The emotive and the descriptive meaning can be represented by two different patterns of argument, argument from values and argument from classification.

#### 3.1 Reasoning from classification

As Zarefsky and Schiappa suggested, naming reality can be conceived as a reasoning process aimed at the attribution of a predicate to a subject. The

1 Al Kamen, The End of the Global War on Terror. *The Washington Post* 24 March 2009, retrieved from [http://voices.washingtonpost.com/44/2009/03/23/the\\_end\\_of\\_the\\_global\\_war\\_on\\_t.html](http://voices.washingtonpost.com/44/2009/03/23/the_end_of_the_global_war_on_t.html) (accessed on 18 March 2012).

most generic description of this mechanism has been introduced by Hastings (1963) and developed by Walton (1996: 54), who provided an abstract structure of argument representing the combination of the rhetorical predicate of “classification” (Hobbs 1979: 68) and the *modus ponens* logical rule. In the following scheme the rhetoric, or rather semantic, relation is stated in a generic fashion, not specifying on which grounds the predicate is attributed to the entity:

PREMISE 1:	For all $x$ , if $x$ has property $F$ , then $x$ can be classified as having property $G$ .
PREMISE 2:	$a$ has property $F$ .
CONCLUSION:	$a$ has property $G$ .

Table 1. Argument scheme classification / *modus ponens*

The generality of the semantic principle risks leading to forms of inference of the kind “ $x$  is blue, therefore  $x$  is a man”. For this reason it needs to be specified taking into consideration the ancient maxims of inference (Stump 1989, Green-Pedersen 1984). The passage from the property stated in the antecedent to the property attributed in the consequent needs to be grounded on the semantic definitory relation (Walton & Macagno 2008), namely the relation concerning the identity and difference between two predicates (Aristotle *Topics* 102a, 5-9). This type of argument can be represented as follows (Walton & Macagno 2010: 39):

PREMISE 1:	For all $x$ , if $x$ fits definition $D$ , and $D$ is the definition of $G$ , then $x$ can be classified as $G$ .
PREMISE 2:	$a$ fits definition $D$ .
CONCLUSION:	$a$ has property $G$ .

Table 2. Argumentation scheme 1: Argument from definition

As Aristotle pointed out, the concept of definition can include different types of equivalences, the most famous (and controversial) of which is the definition by genus and difference. However, in addition to this classical method, the same concept can be defined in various fashions. For instance, “man” can be defined by genus and difference as the “reasonable animate being”, by property as the “being who can learn grammar”, by physical parts as the “being who has a head, two arms, two legs, etc...” The process of classification can be conceived as a type of reasoning proceeding from

definition (man is a reasonable animal; man is a biped without feathers), other definitional propositions (descriptions, operative definitions), or heuristic processes establishing an identity (classification by contraries, analogy, etc.). Moreover, different definitional sentences trigger different types of reasoning. The definition by genus and difference leads to a classificatory (affirmative) conclusion by means of a deductive *modus ponens*. However, definitions by contrary can only classify an entity by denying what the *definiendum* is not, proceeding by *modus tollens* or *modus ponendo tollens*. Other definitions warrant a classification abductively (parts are signs of the entity in definitions by parts; a cause is the explanation of an effect in an operative definition) or by means of analogy. The generic reasoning pattern named “classification” becomes on this view an umbrella term encompassing different types of semantic principles, each of them triggering distinct types of reasoning to support a classificatory conclusion. Below the most important types of defintory premises and the related principle of inference leading to an affirmative conclusion are represented:



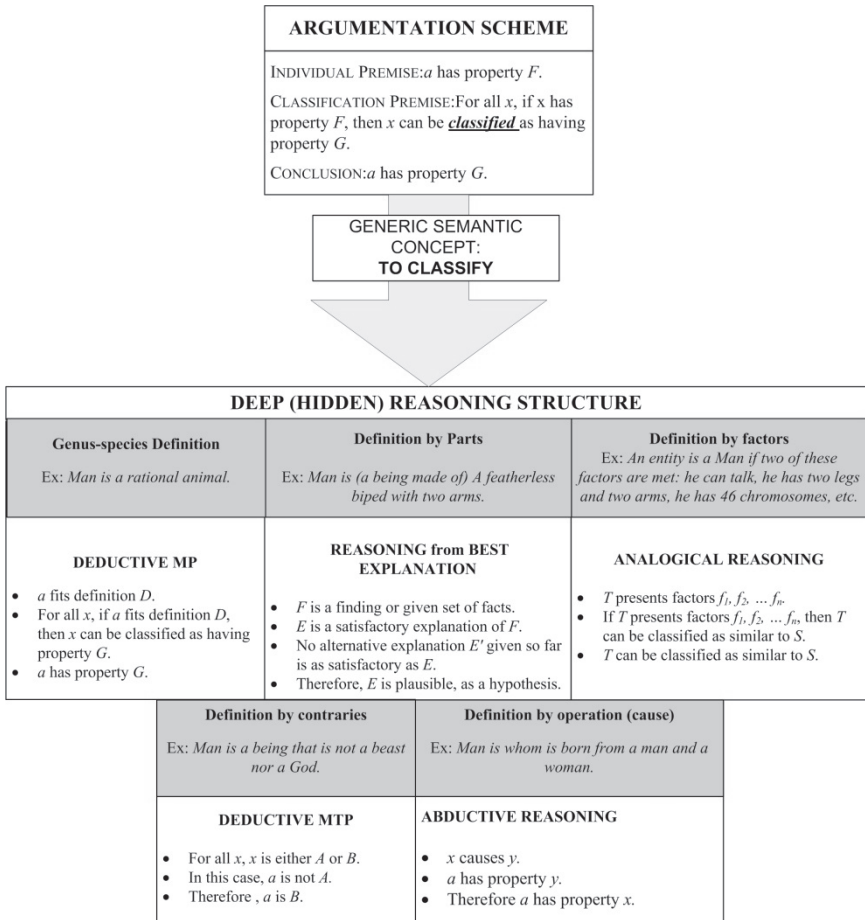


Figure 1. The reasoning and semantic dimensions of argument from classification

As shown in figure 1 above, argument from classification can be considered as a generic pattern describing the “descriptive meaning” of a word, which can be further specified by pointing out the semantic relationship between *definiens* and *definiendum*. However, the semantic relation determines the “deep logic” of the argument: the surface defeasible modus ponens hides more complex types of reasoning based on the meaning of the definitional premise.

### 3.2 *Argumentation from values*

Stevenson described the emotive meaning as a propensity to encourage actions. The relationship between words, meaning and values (or, rather, hierarchies of values, Perelman & Olbrechts-Tyteca 1951), can provide an explanation, from a rhetorical perspective, of the reason why words can lead to value judgments and decisions. On this view, values can be thought of as the reasons for classifying something as desirable or not, and at the same time be used to encourage action. By pointing out the qualities of a course of action, an event or an object that the interlocutor considers as valuable (desirable), the speaker can provide him with a reason to act in a specific fashion. This twofold process of reasoning can be described as follows:  $x$  (an action, an object, or a viewpoint) can be judged positively or negatively according to a value (or rather a hierarchy of values)  $V$ ; according to the desirability of  $x$ ,  $x$  can become an action worthy for the agent or not. Values represent the criterion for establishing the desirability of a course of action, and the generic form of reasoning based on them can be represented as follows (Walton, Reed & Macagno 2008: 321):

PREMISE 1:	Value $V$ is positive (negative) as judged by agent $A$ (value judgment).
PREMISE 2:	The fact that value $V$ is positive (negative) affects the interpretation and therefore the evaluation of goal $G$ of agent $A$ (If value $V$ is good (bad), it supports (does not support) commitment to goal $G$ ).
CONCLUSION:	Value $V$ is a reason for retaining (retracting) commitment to goal $G$ .

*Table 3.* Argumentation scheme 2: Argument from values

This pattern of argument can be further specified considering its two dimensions, the process of evaluation (a specific kind of classification, Von Wright 1963) and the decision-making reasoning. The first step consists in classifying an action or a state of affairs as desirable or not according to our hierarchy of values. For instance, security, justice or richness can be evaluated as preferable to peace and human life by someone, while others can consider these latter values as the most important ones. The classification of an action as “an act of war” can be evaluated differently according to the hierarchies of values of the audience. The different reasons underlying this value judgment were outlined in Aristotle’s *Rhetoric* and *Topics* in form of

commonplaces or topics. Such topics can be conceived as possible different ways of defining what is “good” according to possible situations and points of view. Since the meaning of “good” is partially determined by the object of its predication (Vendler 1964), these commonplaces are useful for warranting the classification, such as the following ones (*Rhetoric* 1363b 13-16):

Now we call “good” what is desirable for its own sake and not for the sake of something else; that at which all things aim; what they would choose if they could acquire understanding and practical wisdom; and which tends to produce or preserve such goods, or is always accompanied by them;

What “is to be chosen for its own sake” can be established on the basis of a person's experiences or culture. On this perspective, hierarchies of values are forms of relativistic definitions of what is desirable.

The second component of reasoning from values is the reasoning passage from moral judgment to action. The relationship between will, and desire, and action is underscored in Aristotle’s *Nicomachean Ethics*. What is good, or appear as such, is maintained to be the goal of a decision to act (*Nicomachean Ethics* 1113a, 15), as “everything aims at the good” (*Topics* 116a, 18). For instance, an act of war can be judged negatively, and for this reason it can be used as a reason for criticizing a military intervention or voting against a party supporting it. The decision-making process can be thought of as a pattern of reasoning connecting an action, or rather a “declaration of intention” or commitment (von Wright 1972: 41) with its grounds (Anscombe 1998: 11). Depending on whether the speaker is assessing a *specific* course of action based on its consequences, or a possible way to achieve a goal, the type of reasoning has different forms. The first and simpler form of argument is the *argument from consequences* (from Walton, Reed & Macagno 2008: 332):

PREMISE 1:	If <i>A</i> is brought about, good (bad) consequences will plausibly occur.
PREMISE 2:	What leads to good (bad) consequences shall be (not) brought about.
CONCLUSION:	Therefore <i>A</i> should be (not) brought about.

Table 4. Argumentation scheme 3: Argument from consequences

For instance, classifying an operation as an “act of peace” or as “pacification” can trigger a reasoning from positive consequences: since such an operation *leads* to peace, and peace is desirable, the operation shall be supported.

The other form of reasoning, called *practical reasoning*, proceeds from a purpose to the possible means that can bring it about (Walton, Reed & Macagno 2008: 323):

PREMISE 1:	I (an agent) have a goal $G$ .
PREMISE 2:	Carrying out this action $A$ is a means to realize $G$ .
CONCLUSION:	Therefore, I ought to (practically speaking) carry out this action $A$ .

*Table 5.* Argumentation scheme 4: Practical reasoning

For instance, freeing people from want and need in countries governed by dictators can be regarded as highly desirable. In order to achieve this goal, the agent can perform various actions that are the sufficient conditions to bring about the desired state of affairs. Among this paradigm of choices, the agent chooses the best one, based on his values or preferences. Obviously, such a paradigm can be altered by the speaker: for example, waging war can be shown to be the only (reasonable, possible) means for freeing “enslaved” populations.

#### 4. Presupposing definitions

Stevenson pointed out how words can be used to affect the interlocutor’s decisions. The distinction between the different types of reasoning triggered by the use of a word can show how redefinitions can affect the evaluation of the state of affairs referred to. Definitions, or definitory statements, are the premises of classificatory reasoning, which are often taken for granted because considered as part of the common ground. However, words can be, and often are, redefined. On the one hand, redefinition is not only a common move, but it is often necessary in order to clarify concepts or highlight new perspectives on them. On the other hand, however, by modifying the definition of a word the speaker can ground the implicit evaluative reasoning of the interlocutor on different premises. How is it possible to distinguish mischievous uses of definition from the legitimate or persuasive ones?

#### 4.1 Definitions and implicit definitions

Definitions can be considered as statements concerning the identity between two concepts, which can be shared or controversial. When a new definition is advanced, it becomes a standpoint that needs to be supported by reasons if not accepted by the interlocutor. A definition, or a redefinition, is an implicit claim in favour of a new use of an existing word (Schiappa 2003), and needs to be open to challenge. We can conceive a redefinition as a standpoint conflicting with the shared opinion on a word use and for this reason it is presumed not to be accepted. There is nothing wrong with redefining a word; the crucial problem is how a redefinition is introduced. For instance, we can consider how Obama redefined the concept of “hostilities” to classify American airstrikes in Libya. In order to avoid Congress authorization to continue the “hostilities”, Obama adapted the meaning of such word to exclude bombings and operations conducted by unmanned aircrafts (*Obama Administration letter to Congress justifying Libya engagement*, June 15th, 2011, p. 25):

Implicit redefinition: “Hostilities”

The President is of the view that the current U.S. military operations in Libya are consistent with the War Powers Resolution and do not under that law require further congressional authorization, because U.S. military operations are distinct from the kind of “hostilities” contemplated by the Resolution’s 60 day termination provision. [...] U.S. operations do not involve sustained fighting or active exchanges of fire with hostile forces, nor do they involve the presence of U.S. ground troops, U.S. casualties or a serious threat thereof, or any significant chance of escalation into a conflict characterized by those factors.

Instead of explicitly arguing for a new definition, or rather a specification, of the concept, Obama takes it for granted. He does not reject the shared one or attacks it based on its vagueness. He does not even suggest that it should be better clarified. Instead, he supports the claim that the US are not engaged in any hostilities in Libya as ground troops have not been deployed, nor have ground battles been fought. He takes for granted that “hostilities” means only “active fighting by ground troops”, which does not correspond to any accepted definition of the term according to ordinary or military dictionaries. This move suggests a crucial question: How is it possible to take for granted a proposition, and what are the boundaries of this implicit (non-)act of

discourse? The notions of pragmatic presupposition and act of presupposing can provide a possible explanation.

#### 4.2 *Pragmatic presuppositions*

Presuppositions are considered as properties of the use of sentences, or rather, statements (Strawson 1950, 1952, Karttunen 1973, Kempson 1975, Wilson 1975, Keenan 1971). This pragmatic view extends the notion of presupposition to several phenomena of meaningfulness constraints (Austin 1962: 34; 51), such as selectional restrictions, coherence relations and felicity conditions. Several phenomena are labeled as presuppositions, including some dimensions of the representation of lexical meaning, the wider class of felicity conditions of speech acts and coherence relations. The common characteristic of all these phenomena is that a proposition  $p$  is presupposed when it is taken for granted in performing a speech act, whose felicity depends on the interlocutor's acceptance of  $p$ . To presuppose a proposition is "to take its truth for granted, and to assume that others involved in the context do the same" (Stalnaker 1970: 279, Stalnaker 1974: 200). This "taking a proposition for granted" has been analyzed as Stalnaker as a propositional attitude, which can be interpreted as an action of a kind (Stalnaker 2002: 701). As Kempson put it (1975: 190), presupposing amounts to treating a proposition as part of the common ground:

The speaker believes that the hearer knows (and knows that the speaker knows) a certain body of propositions (i.e. there is a Pragmatic Universe of Discourse) and in making a certain utterance " $*\psi p$ " he believes that the hearer, knowing the conventions of the language and hence the conditions for the truth of the proposition in question, will recognise a subset of those conditions as being part of that Pragmatic Universe of Discourse and hence neither assertible, deniable or queriable [...].

In particular, for the purpose of this paper a specific type of pragmatic presupposition will be inquired into, the presupposition of definitional sentences. For this reason, it is necessary to investigate how definitions can be presupposed in discourse, or rather how they can be triggered.

### 4.3 *Presuppositions of discourse relations*

Definitions, being the implicit premises of a classificatory argument, need to be inquired into taking into consideration the linguistic structure of discourse relations, or, rather, connectives. Karttunen (1973: 176) described how presuppositions can be triggered by predicates of higher level, the connectives, whose linguistic arguments are discourse sequences. Connectives link sequences and presuppose specific relations between them. For instance, we can consider the following famous case (Lakoff 1971: 133):

I. John is tall, but he is no good at basketball.

Lakoff notices that (I) is composed of an assertion (John is tall, and he is no good at basketball) and a presupposition (If someone is tall, then one would expect him to be good at basketball). The effect is a denial of expectation, which was described by Ducrot as the contradiction by the second conjunct of a presupposed conclusion (in this case, “John is good at basketball”) (Ducrot 1978). Similarly, the connective “and” presupposes a common relevance or topic (Lakoff 1971: 128, Kempson 1975: 58), which can be observed in the following cases (Kempson 1975: 56; 61):

II. The Lone Ranger rode off into the sunset and mounted his horse.

III. Pope John is dying and the cat is in the bath.

Both sentences are unsound because the relationship between the two conjuncts seems to be missing, or rather is unavailable to the interlocutors in normal conditions. In (II) the conjunction presupposes a temporal sequence that is commonly perceived as impossible, while in (III) the (causal) relationship cannot be even retrieved. Subordinate connective specify more precisely the type of relationship between the sequences. For instance, the predicate “therefore” presupposes that the first sequence is a reason supporting the second one (see also Grice 1975: 44).

Both in case of coordination and subordination, text sequences are connected in a similar fashion. In subordination the predicate is explicit and imposes a set of specific coherence conditions, or pragmatic presuppositions (Vanderveken 2002: 47, Bach 2003: 163), on its arguments (Grimes 1975: 162). In coordination, an explicit or implicit predicate hides a deeper relationship (Ballard, Conrad & Longacre 1971) that needs to be reconstructed in order to understand the role and the conditions of the discourse segments or sequences. For instance, coordination can express

temporal, causal, explanation relations, imposing specific requirements on their sequences, such as a causal or temporal order of the sequences. In all cases, a “high level notion” or “logical-semantic connective” (Rigotti 1993, Rigotti 2005, Rigotti & Rocci 2001) connects the propositions expressed by the clause; such a notion, or predicate, can be expressed or not, and specified or not. In all cases, the sentences or clauses are connected by an abstract, high level and generic semantic relation that imposes specific requirements on its arguments. There can be several high-level relations: explanation, narration, contrast, etc. (see Lascarides & Asher 1993). However, only one of such relations, namely motivation or support, will be considered here. We can analyse the following interpretation and reconstruction of the aforementioned argument used by Obama to classify the airstrikes in Libya:

- IV. (A) Our operations do not involve the presence of U.S. ground troops. (B) (*therefore*) Our operations are not “hostilities”.

In this case, a higher level predicate connecting the discourse moves needs to be reconstructed. We can represent it linguistically as the connective “therefore”, expressing a relation of motivation (Rigotti & Rocci 2006). Such a relation needs to support the attribution of a predicate (to be a case of hostility) in *B* to the same subject of the previous sequence (*A*). The attribution of a predicate on the basis of actions or qualities attributed to the same subject can be usually presumed to be a classification. Obviously, the specification of the relation depends on several factors, such as the type of property attributed. This relation requires that the quality or event expressed in the first sentence represents a classificatory, or definitional, principle for the attribution of the quality in the second sentence (Kempson 1975: 109-110). In this specific case, the fact, event or quality need to instantiate a definition, or definitional principle, of “to be a case of hostilities”. We can represent the structure of the presuppositions as follows:



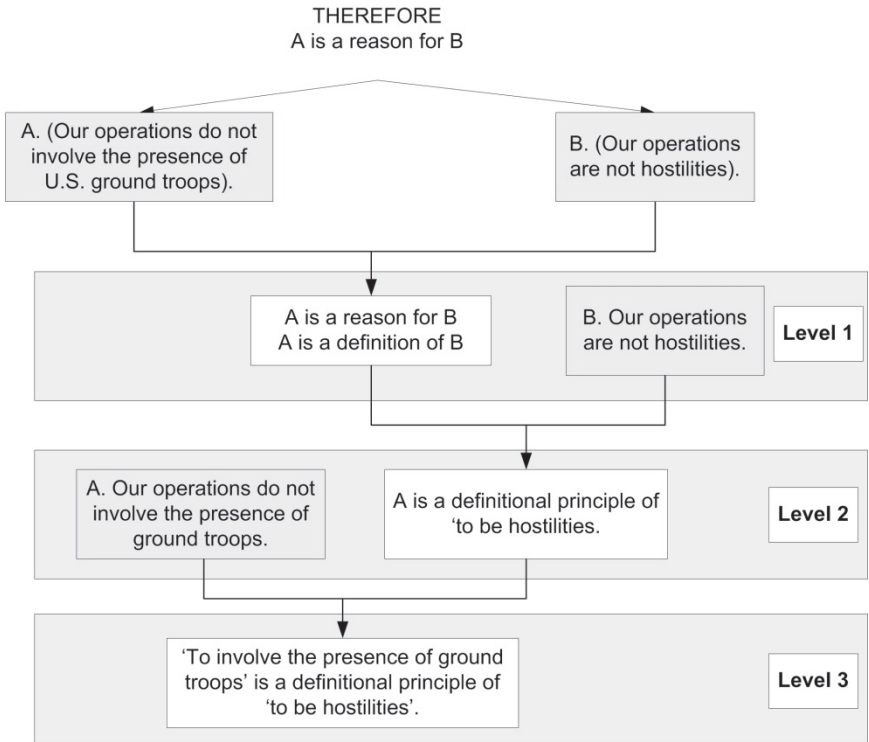


Figure 2. Presuppositions of “therefore”

The abstract relation of coherence (Asher & Lascarides 2003, chap. 7) is further specified according to the three levels of reasoning performed by the addressee for reconstructing (or rather justifying, see Kamp 2001) the presupposition. From the meaning of the connector and the information of sequence B (level 1), a first specification of the discourse relation is drawn (A is a definition of “to be hostilities”). The combination of this first conclusion with A (level 2) leads to the final reconstruction, which indicates the actual relation between the two sequences (“to involve the presence of ground troops” is a definitional principle of “hostilities”).

## 5. The act of presupposing

As seen above, from a linguistic perspective the presuppositions of connectives are requirements for the coherence of a text or discourse. The use of such requirements leads us to the other crucial perspective, the pragmatic one.

### 5.1 *Presupposition as an implicit act*

From the accounts of pragmatic presupposition mentioned at 4.1 above, two crucial elements emerge: 1. Presupposition can be considered as a decision to treat a proposition as shared; 2. Presuppositions are crucially related to the speaker and hearer's beliefs and knowledge (Schwartz 1977: 248). The definition of a linguistic phenomenon in terms of beliefs or assumptions risks confounding the phenomenon with its accidental effects or possible explanations. How can a speaker believe or assume that a proposition is shared by the hearer? How would it be possible to presuppose propositions that are known not to be shared, without the sentence being infelicitous? A possible explanation consists in analyzing the pragmatic presuppositions as an act consisting in treating a proposition as shared, and investigating its conditions and essential requirements.

Ducrot pointed out the strict relationship between a speech act and its conditions. He noticed that the performance of a speech act amounts to implicitly performing a hidden, or rather implicit, act: presupposing (Ducrot 1968: 87). For instance, by asserting that, "We are freeing the people of Afghanistan from need" the speaker is deploying a dialogical world in which people in Afghanistan are in need and need is a form of burden. Only in such a world his statement is felicitous. On Ducrot's view, by presupposing the speaker modifies the dialogical situation, and set the boundaries of the interlocutor's future actions (Ducrot 1972)<sup>2</sup>, or rather the conditions for the

2 On Ducrot's view, the communicative game resembles a chess game, in which the possibilities are set by means of presuppositions: "dans ce combat simulé –qui substitue aux possibilités réelles, dues à la force, les possibilités morales dues aux conventions- les règles permettent aux joueurs de se contraindre mutuellement à certaines actions, et de s'en interdire certaines autres" (Ducrot 1968: 83, 1972: 27).

continuation of the future dialogue game (Ducrot 1991: 91). Not accepting a presupposition amounts to ending the dialogue, something like knocking over the chessboard.

This perspective takes into account solely the structure of the dialogue move, and not its possibility. A speech act of the kind “Bob’s brother is feeling bad today” would fail to fulfil the purpose of informing the hearer if the latter knows that Bob is an only child, or if he does not know Bob at all. In order to account for the effect of a move, and therefore its possibility conditions and fallacious uses, it is necessary to take into consideration the relationship between the speaker and the hearer’s knowledge. This relation can be examined starting from a case studied by Ducrot (1966: 42). He considered an imaginary conversation between the enemies of Cesar or Napoleon during the Roman consulate or the French Republic. In this conversation, they talk about “the magnificence, or the richness or the wisdom of the King”. In this case the speakers presuppose false or unshared propositions (“Cesar – or Napoleon – is a king”). However, their assertions, far from being void, might have caused them serious troubles for their meaning. This case illustrates a crucial problem of presuppositions, the possibility of treating as shared an unshared proposition, relying on the hearer’s capacity of reconstructing, or rather “accommodating” it (Lewis 1979; Von Stechow 2008). From the analysis of the limits of such a process of reconstruction it is possible to understand the conditions characterizing the speech act of presupposition.

## *5.2 The limits of presupposing*

Presuppositions, on Ducrot’s view, need to be accepted in order for the dialogue to be possible. However, at the same time presuppositions need to be known by the interlocutor. From a pragmatic perspective, the possibility of presupposing information not shared, or not known to be shared, needs to be accounted for. On Lewis’ perspective (Lewis 1979), the hearer reconstructs the presupposed and not shared propositions in order to avoid communicative failure (Von Stechow 2008); in other words, he accommodates the missing and necessary information (Lewis 1979: 340):

If at time  $t$  something is said that requires presupposition  $P$  to be acceptable and if  $P$  is not presupposed just before  $t$ , then – *ceteris paribus* and within certain limits – presupposition  $P$  comes into existence at  $t$ .

The crucial problem of this view is to determine how a presupposition can come into existence and be added to the shared propositions. On Soames' view, accommodation is possible when no objections are raised, namely the interlocutor has already accepted the proposition (it is part of the common ground) or it is not conflicting with it (Soames 1982: 486):

Utterance Presupposition An utterance  $U$  presupposes  $P$  (at  $t$ ) iff one can reasonably infer from  $U$  that the speaker  $S$  accepts  $P$  and regards it as uncontroversial, either because

- a.  $S$  thinks that it is already part of the conversational context at  $t$ , or because
- b.  $S$  thinks that the audience is prepared to add it, without objection, to the context against which  $U$  is evaluated.

Soames explains the phenomenon of accommodation in terms of the speaker's beliefs regarding the interlocutor's common knowledge. However, how is it possible to evaluate a belief? Is presupposition dependent on personal beliefs?

A possible answer is suggested by Stalnaker (1998). He explains the relationship between speaker's and hearer's knowledge in the process of accommodation as a *presumption* of the speaker that the presupposed information is *available* to his or her audience (Stalnaker 1998: 8). The speaker acts holding the conclusion of his or her presumptive reasoning as true until contrary evidence is provided. For instance we can consider the following variants of the statement made by Obama before the Congress and analysed in figure 2 above:

- A. Our intervention cannot be considered as hostilities. We have not used weapons or the military.
- B. Our intervention is not a *çatışma*. They have not intervened.
- C. Our intervention cannot be considered as hostilities. It is fast and well done.
- D. Our intervention cannot be considered as hostilities. It does not involve cooking of potatoes.

These four cases differ for different reasons. In (A), the speaker grounds his presupposition on the fact that people (and congressmen) usually know what "hostilities" are, and that "using weapons or the military" is a possible

criterion for classifying actions as hostilities. In (B), however, it is impossible to reconstruct and accept the presupposition, as an essential requirement clearly fails for two reasons. The speaker cannot presume that North American congressmen know the meaning of a Turkish word, “*çatışma*”. Moreover, since no information has been provided on the entity to which “they” refers. Such presuppositions (the definition of “*çatışma*” and the referent of “they”) cannot be accommodated, cannot be reconstructed, as they are not “rhetorically bound to the context” (Asher & Lascarides 1998: 277), nor they are related to propositions presumed to be known. In this case, the process of reconstruction shown in figure 2 can fail at level 1 or 2, as the speaker may not understand the meaning of the sequences connected and therefore retrieve their relationship, or he can understand their relationship but cannot reconstruct the definitory statement. Reconstruction is not the only process which needs to be considered for analysing presuppositions, as (B) does not represent the only case in which the speech act is infelicitous because of presuppositional failure. In (C) and (D) the hearer can understand the nature of the proposition taken for granted (a definitory statement) and connect it with his or her background knowledge. However, in (C) the hearer cannot accept that the property of “being nice and well done” is a definition of an action (hostilities). In this case, the process of presupposition reconstruction represented in figure 2 above fails at level 2. In (D), the presupposition can be reconstructed and its nature of definitory statement accepted. However, no congressmen and presumably no English speaker can accept that “cooking of potatoes” is a definition of “hostilities”. The conclusion of the process of reconstruction outlined in figure 2 above cannot be accepted and fails at level 3. The *possibility* of presupposing needs therefore to be distinguished from the *acceptability* of a proposition taken for granted. By distinguishing the two dimensions of accommodation it is possible to distinguish between four different cases:

- i) the presupposition can be reconstructed and accepted as a background assumption (case A);
- ii) the presupposition cannot be reconstructed (case B);
- iii) the presupposition can be reconstructed but its function (nature, structure) cannot be accepted (case C);
- iv) the presupposition can be reconstructed but its content cannot be accepted (case D).

These possibilities allow us to outline the possible felicity conditions of the implicit speech act of presupposing, building on Austin's and Searle and Vanderveken's accounts of speech act conditions (Austin 1962: 14-15, Searle & Vanderveken 1985: 13-19, Holdgraves 2008: 13):

<i>Essential Condition:</i>	Speaker ( <i>S</i> ) sets the presupposed proposition ( <i>pp</i> ) as a condition of the felicity of his speech act ( <i>SA</i> ); if Hearer ( <i>H</i> ) does not accept <i>pp</i> , <i>SA</i> will be void.
<i>Propositional Condition:</i>	<i>pp</i> is a proposition that can be reconstructed by <i>H</i> .
<i>Preparatory Condition:</i>	<i>S</i> can presume that <i>H</i> can reconstruct and accept <i>pp</i> .
<i>Sincerity Condition:</i>	<i>S</i> believes that <i>pp</i> ; <i>S</i> believes that <i>H</i> can reconstruct and know or accept <i>pp</i> .

Table 6. Felicity conditions for the speech act of presupposing.

This speech act has a direction of fit from World (of the Hearer) to Words (of the Speaker), and its goal is to set the propositions that the hearer needs to accept for the dialogue to continue. The possibility of reconstructing the presupposition is indicated as a propositional condition: *H* needs to be able to draw *pp* from the linguistic and pragmatic elements provided. The acceptability of the presupposition is governed by both the preparatory and the sincerity condition. The sincerity condition expresses the conditions that the tradition on pragmatic presupposition considered as essential, while the preparatory condition, framed as a presumption, is aimed at bridging the gap between the speaker's and hearer's mind from an epistemic and argumentative perspective, without resorting to the psychological notion of belief.

This treatment of presupposition as a kind of implicit speech act can explain also the particular types of moves in which the speaker takes for granted a proposition known to be false or unknown by the hearer, such as the cases of Napoleon and Caesar mentioned above. Here the speaker can presume and believes that the hearer can reconstruct the presupposition "Caesar (or Napoleon) is a king", but at the same time he presumes and believes that he cannot know or accept it, as it is false. Ducrot described this phenomenon as a form of connotation, in which the utterance becomes a sign aimed at communicating the conditions of its use (Ducrot 1968: 44). In a speech act perspective, this particular use of presupposition can be regarded as an indirect speech act, where the act setting out the conditions of a move needs to be interpreted as a type of assertive (Hickey 1993: 107).

## 6. Presuppositions as presumptive reasoning

The most important aspect of the speech act of presupposition is the preparatory condition, stating that the speaker can presume that the hearer can reconstruct and accept (or in a stronger sense, know) the proposition to be presupposed. This condition sets out the grounds of the *reasonableness* of speaker's presupposition, and tries to provide a possible answer to the following crucial question: Why and how can a speaker presuppose a proposition? The concepts of speaker's "belief" or "thinking" mentioned in the theories on pragmatic presupposition ascribe the phenomenon of presupposition to internal cognitive processes. However, such explanations cannot provide criteria for distinguishing between reasonable uses of presupposition from absurd or manipulative ones. If we examine presupposition in terms of presumptive reasoning, we can analyze its reasonableness by assessing the reasons supporting its fundamental requirement, the fact that the presupposed proposition *can* be shared. This concept (partially hinted at by Strawson, see Strawson 1971: 58-59, Kempson 1975: 166-167), shifts the traditional psychological explanation onto an epistemic level.

### 6.1 Presumptive reasoning

Presuppositions can be conceived as the conclusion of presumptive reasoning. The speaker cannot know the other's mind, but only advance a tentative and defeasible conclusion based on a form of reasoning in lack of evidence (Rescher 1977: 1). He or she draws specific conclusions on the other's mind based on general principles such as "Speakers belonging to a specific speech community usually know the meaning of the most important words of the language used therein". In other words, he is only presuming such knowledge.

Presumptions cannot *prove* a conclusion; they rather intervene when it is *not possible* to demonstrate a conclusion (Blackstone 1769: 371). This type of reasoning is rebuttable and defeasible (Hart 1961: 10), as its characteristic consists in supporting a conclusion until contrary evidence is produced. However, its inherent defeasibility has a fundamental effect on the dialogical

setting, the shifting of the burden of producing evidence (or proving a proposition) onto the other party. For instance, the fundamental legal presumption is the innocence of the defendant. This does not mean that the defendant is innocent, but simply that he is considered as such until he is proved guilty (beyond a specific standard of proof). The other party, the prosecution (or in civil cases the plaintiff) has to provide evidence to rebut this presumptive conclusion.

The legal framework provides a general idea of the structure of this reasoning in everyday argumentation. Presumptions work to move the dialogue further when knowledge is lacking. Their role is to shift the burden of proof onto the other party, who can reject the proposition only by providing contrary arguments or positive facts leading to a contrary conclusion. If not rebutted, the speaker can consider it as tentatively proved, and move the dialogue further. Rescher outlined the structure of this type of inference as follows (Rescher 2006: 33):

RULE	$P$ (the proposition representing the presumption) obtains whenever the condition $C$ obtains unless and until the standard default proviso $D$ (to the effect that countervailing evidence is at hand) obtains.
FACT	Condition $C$ obtains.
EXCEPTION	Proviso $D$ does not obtain.
CONCLUSION:	$P$ obtains.

Table 7. Structure of presumptive reasoning

The *Rule* of presumption links the acceptability of a proposition  $P$  (for instance, the defendant is innocent) to a condition  $C$  (for instance, he denies the crime he is charged with) until a specific default proviso  $D$  obtains (for instance, he is found guilty beyond reasonable doubt). If he denies the charge and is not found guilty beyond reasonable doubt, he is to be found innocent. This type of reasoning can be applied to the analysis of the conditions of presuppositions to assess when and whether the speaker can reasonably take a proposition for granted. This pattern of reasoning outlines the structure of the reasoning underlying his “belief” or “thinking” that the interlocutor accepts or knows the presupposed proposition.



## 6.2 Presumptions and redefinitions

The structure of presumptive reasoning mentioned above can be applied to the cases of redefinition cited, and in particular the persuasive definitions of “hostilities” and “peace”. In the first case, Obama took advantage of the absence of an explicit definition in the War Powers Resolution. However, the absence of an explicit definition does not amount to the absence of a shared one and, therefore, it cannot result in the acceptability of *any* definition. We can reconstruct Obama’s reasoning as follows:

Common knowledge	The accepted meaning of ‘hostilities’ is «overt act of warfare».
RULE	<i>The congressmen should know (be committed to) the meaning of “hostilities” (P) whenever such a word is used with its commonly accepted meaning, or when the speaker redefined it supporting it by reasons (C) (unless the interlocutor does not master the language, belongs to a different culture or community, etc.). (D)</i>
FACT	The commonly accepted definition of ‘hostilities’ is «overt act of warfare». (C)
EXCEPTION	It is not the case that the audience does not know the language or belongs to a different community of speaker (or culture). (non-D)
CONCLUSION:	<i>The audience should know that ‘hostilities’ means «presence of land troops and sustained fighting». (P)</i>

Table 8. Reconstruction of Obama’s redefinition of ‘hostilities’.

In this case, the crucial problem was not the absence of a definition of “hostilities”, but its contrary, the presupposition made by Obama of its existence and sharedness. Obama takes for granted not only that a specific definition exists, but also that such a definition conflicting with the common knowledge is accepted by everybody.<sup>3</sup> The mischievous nature of the move lies in presenting as accepted a proposition incompatible with the accepted one, shifting the burden of disproving it onto the addressees.

The effect of a presumption of meaning is much greater when a concept is “essentially contested” (Gallie 1956). Concepts such as “art”, “freedom”, “peace” or “democracy” are vague and controversial, and admit of borderline cases that cannot be clearly classified. For these reasons, there can be

3 See the commonly accepted definition of ‘hostilities’ at <http://www.merriam-webster.com/dictionary/hostilities> (last retrieved on 3 December 2012).

different definitions for the same concept. “Peace” belongs to this category of controversial and contested concepts. However, usually all definitions (etymological, by description, by qualitative parts) share one fundamental generic feature, absence of conflict. If we analyze Obama’s redefinition of “peace” in his Nobel Prize address, we can notice that presupposes a redefinition by means of a twofold move. First, he introduces a difference of the commonly accepted genus “absence of visible conflict”, stating that it shall “be based on inherent rights and dignity [...] economic security and opportunity”. Then, he implicitly replaces the genus, without providing any reason, claiming that “true peace is *not just freedom from fear*, but freedom from want”. He presupposes that the second sequence replaces the accepted meaning; however, the very definition he is arguing against presupposes a new genus for peace, “freedom”, instead of “absence of conflict”. The underlying presumptive reasoning proceeds as follows:

Common knowledge	The accepted meaning of ‘peace’ is «absence of conflict».
RULE	<i>The interlocutor should know (be committed to) the meaning of ‘peace’ (P) whenever such a word is used with its commonly accepted meaning, or when the speaker redefined it supporting it by reasons (C) (unless the interlocutor does not master the language, belongs to a different culture or community, etc.). (D)</i>
FACT	The commonly accepted definition of ‘peace’ is «absence of conflict»; the contextually accepted one is its partial redefinition by the speaker: «absence of conflict based rights and opportunities (based on freedom from want)...». (C)
EXCEPTION	It is not the case that the audience does not know the language or belongs to a different community of speaker (or culture). (non-D)
CONCLUSION:	<i>The audience should know that ‘peace’ means ‘a kind of freedom.’ (P).</i>

Table 9. Reconstruction of Obama’s redefinition of ‘peace’.

The conclusion of the presumptive reasoning does not follow from the premises. In fact, it actually contradicts them. Obama redefines an essentially controversial concept and shifts the burden of proof onto the audience. By presuming the acceptability of his definition, he leads a possible opponent to disprove it by providing a more accepted alternative, which cannot be easily done.

The analysis of presupposition as presumptive meaning shows the boundaries of implicit redefinition (or rather persuasive definition) in terms

of reasonableness of the presumptive reasoning on which their implicit dimension is based. Moreover, the description of presuppositions as presumptions underscores another crucial effect of implicit redefinitions, the dialectical effect of shifting the burden of proof. Following Walton and Krabbe's dialectical models (1995), we can represent a verbal exchange as an alteration of the agents' commitment store, which contains all the statements that the participant has conceded or accepted during the course of the dialogue. In a dialogue not all the commitments are explicit. The interlocutors can interact because they share the definitions of the words used, the rules of the dialogue, procedures and encyclopedic knowledge regarding the place where they are. Some of these "dark-side" commitments (Walton & Krabbe 1995: 11) are the outcome of previous dialogues, and represent the propositions that the interlocutors have accepted or stated. In a dialogue, presuppositions are implicit activations of dark side commitments (see Corblin 2002): they refer to propositions already accepted by the parties to move the commitments further. Presupposing unshared propositions is a twofold dialectical strategy. On the one hand, presuppositions are commitments: presupposing an unshared proposition means committing the hearer to a view that he or she never accepted, and that has to be denied in order to be deleted from the commitment store. On the other hand, presuppositions are the conclusions of implicit presumptive reasoning, and therefore their denial needs to be supported by an argument that rebuts the presumption.

## 7. Conclusion

Implicit redefinitions can be crucial and dangerous instruments of persuasion and manipulation. Stevenson underscored how they can be used to redirect emotions and affect judgments and decisions. By modifying the meaning of a word that triggers positive or negative judgments, the speaker can influence the hearer's perception and evaluation of a state of affairs, and alter his course of action. However, on the one hand redefinitions are not inherently deceptive or fallacious; on the contrary, since they are often necessary. On the other hand, since there are no unique, immutable and universally shared definitions, the risks of definitional relativism and complete freedom of

redefining and manipulating concepts seem inevitable. In this paper the problem of redefinition is investigated from an argumentative and pragmatic perspective.

Redefinitions are analyzed as condensed arguments advanced by explicit and implicit speech acts. Like any other viewpoint or premise in an argument, it needs to be open to criticism if not shared. Implicit redefinitions can be considered as strategies to take a controversial proposition for granted, treating a definition that cannot be agreed upon as shared. Presupposition has been investigated as a form of speech act grounded on the fundamental conditions that the presupposed proposition needs to be *possibly reconstructed* by and *acceptable* to the hearer. However, how is it possible to know the other's mind? How is it possible to know that the interlocutor can retrieve and accept a proposition?

The possible answer suggested in this paper is grounded on the idea of assessing presuppositions taking into consideration the reasonableness of the presumptive reasoning on which they are based. By presupposing, the speaker acts on a reasonable and justified guess on the interlocutor's knowledge. The reasonableness of his act depends on reasoning in lack of evidence, based on what is commonly considered to be the case. Interpreting presuppositions as forms of presumptions, we can draw a line between reasonable and mischievous uses of implicit definitions. There is nothing wrong in taking for granted a definition presumably accepted. However, the act of presupposing a redefinition, namely a description of the meaning of a concept that is *known* to be *non-shared*, amounts to ground the discourse move on a pragmatic contradiction. A new and unshared definition is presented and treated as commonly accepted. This move shifts the burden of proof onto the hearer, who needs to rebut a viewpoint that has never been supported by arguments.

## References

- Anscombe, G.E.M. (1998): "Practical Inference", in Hursthouse R. *et al.* (eds), *Virtues and Reasons*. Clarendon Press, Oxford, 1-34.
- Aristotle (1969): *Topica*, in Ross, W. (ed.), *The Works of Aristotle*. Oxford University Press, Oxford.
- (1984): *Rhetorica*, in Barnes, J. (ed.), *The Works of Aristotle*. Princeton University Press, Princeton.

- (1995): *Nicomachean Ethics*, in Barnes, J. (ed.), *The complete works of Aristotle*, vol. II. Princeton University Press, Princeton.
- Asher, N. & Lascarides, A. (1998): “The Semantics and Pragmatics of Presupposition”, *Journal of Semantics* 15, 239-299.
- Austin, J. (1962): *How to Do Things With Words*. Clarendon, Oxford.
- Bach, K. (2003): “Speech acts and pragmatics”, in Devitt, M. & Hanley, R. (eds), *Blackwell Guide to the Philosophy of Language*. Blackwell, Oxford, 147-167.
- Bierce, A. (2000): *The unabridged Devil’s dictionary*. University of Georgia Press, Athens.
- Blackstone, W. (1769). *Commentaries on the Laws of England. Vol. IV*. Clarendon Press, Oxford.
- Ducrot, O. (1966): “‘Le roi de France est sage’. Implication logique et présupposition linguistique”, *Etudes de linguistique appliquée* 4, 39-47.
- (1968): “Le structuralisme en linguistique”, in Ducrot, O. et al. (eds), *Qu’est-ce que le structuralisme?* Seuil, Paris, 13-96.
- (1972): “De Saussure à la philosophie du langage”, in Searle, J., *Les actes de langage*. Hermann, Paris, 7-34.
- (1978): “Deux mais”, *Cahiers de linguistique* 8, 109-120.
- (1991): *Dire et ne pas dire*. Hermann, Paris.
- Fintel, K. von (2008): “What is Presupposition Accommodation, Again?”, *Philosophical Perspectives* 22, 137-170.
- Gallie, W. (1956): “Essentially Contested Concepts”, *Proceedings of the Aristotelian Society* 56, 167-198.
- Green-Pedersen, N. (1984): *The Tradition of the Topics in the Middle Ages*. Philosophia Verlag, Munich.
- Grice, P. (1975): “Logic and Conversation”, in Cole, P. & Morgan, J. (eds), *Syntax and semantics 3: Speech acts*. Academic Press, New York, 41-58.
- Grimes, J. (1975): *The thread of discourse*. Mouton, The Hague.
- Hart, H. (1961): *The Concept of Law*. Oxford University Press, Oxford.
- Hastings, A. (1963): *A Reformulation of the Modes of Reasoning in Argumentation*. Ph.D. Dissertation, Northwestern University, Evanston.
- Hickey, L. (1993): “Presupposition under cross-examination”, *International Journal for the Semiotics of law* 6 (16), 89-109.
- Hobbs, J. (1979): “Coherence and coreference”, *Cognitive science* 3, 67-90.
- Holdgraves, T. (2008): *Language as social action*. Lawrence Erlbaum, Mahwah.
- Kamp, H. (2001): “Presupposition Computation and Presupposition Justification”, in Bras, M. & Vieu, L. (eds), *Pragmatic and Semantic Issues in Discourse and Dialogue*. Elsevier, Amsterdam, 57-84.
- Karttunen, L. (1973): “Presuppositions of Compound Sentences”, *Linguistic Inquiry* 4 (2), 169-193.
- Keenan, E. (1971): “Two types of presupposition in natural language”, in Fillmore, C. & Langendoen, D. (eds), *Studies in Linguistic Semantics*. Holt, New York, 45-54.
- Kempson, R. (1975): *Presupposition and the delimitations of semantics*. Cambridge University Press, Cambridge.
- Lakoff, G. (1971): “On Generative Semantics”, in Steinberg, D. & Jakobovits, L. (eds), *Semantics: An Interdisciplinary Reader in Philosophy, Linguistics and Psychology*. Cambridge University Press, Cambridge, 232-296.

- Lascarides, A. & Asher, N. (1993): "Temporal Interpretation, Discourse Relations and Commonsense Entailment", *Linguistics and Philosophy* 16 (5), 437-493.
- Lewis, D. (1979). "Scorekeeping in a Language Game", *Journal of Philosophical Logic* 8, 339-359.
- Moliere (2000): "Don Juan", in Wood, J. & Coward, D. (eds), *Molière. The Miser and other plays*. Penguin Books, Toronto.
- Perelman, C. & Olbrechts-Tyteca, L. (1951): "Act and Person in Argument", *Ethics* 61 (4), 251-269.
- Rescher, N. (1977): *Dialectics: a controversy-oriented approach to the theory of knowledge*. State University of New York Press, Albany.
- (2006): *Presumption and the practices of tentative cognition*. Cambridge University Press, Cambridge.
- Rigotti, E. (1993): "La sequenza testuale: definizione e procedimenti di analisi con esemplificazioni in lingue diverse", *L'analisi linguistica e letteraria* 1 (1), 43-148.
- (1995): "Verità e Persuasione", *Il nuovo areopago* 14 (1), 3-14.
- (2005): "Congruity theory and argumentation", *Studies in Communication Sciences* (special issue), 75-96.
- Rigotti, E. & Rocci, A. (2001): "Sens – non-sens – contresens", *Studies in Communication Sciences* 2, 45-80.
- (2006): "Tema-rema e connettivo: la congruità semantico-pragmatica del testo", in Gobber, G., Gatti, M.C. & Cigada, S. (eds), *Syndesmoi: connettivi nella realtà dei testi*. Vita e Pensiero, Milano, 3-44.
- Sager, J. (2000): *Essays on Definition*. John Benjamins, Amsterdam.
- Schiappa, E. (2003): *Defining Reality. Definitions and the Politics of Meaning*. Southern Illinois University Press, Carbondale.
- Schwarz, D. (1977): "On pragmatic presupposition", *Linguistics and Philosophy* 1 (2), 247-257.
- Searle, J. & Vanderveken, D. (1985): *Foundations of Illocutionary Logic*. Cambridge University Press, Cambridge.
- Soames, S. (1982): "How Presuppositions Are Inherited: A Solution to the Projection Problem", *Linguistic Inquiry* 13 (3), 483-545.
- Sorensen, R. (1991): "Vagueness and the Desiderata for Definition", in Fetzer, J. et al. (eds), *Definitions and Definability*. Kluwer, Dordrecht, 71-109.
- Stalnaker, R. (1970): "Pragmatics", *Synthese* 22 (1-2), 272-289.
- (1974): "Presuppositions", in Munitz, M. & Unger, P. (eds), *Semantics and Philosophy*. New York University Press, New York, 197-214.
- (1998): "On the Representation of Context", *Journal of Logic, Language, and Information* 7 (1), 3-19.
- (2002): "Common ground", *Linguistics and Philosophy* 25, 701-721.
- Stevenson, C. (1937): "The Emotive Meaning of Ethical Terms", *Mind* 46, 14-31.
- (1938a): "Persuasive Definitions", *Mind* 47, 331-350.
- (1938b): "Ethical Judgments and Avoidability", *Mind* 47 (185), 45-57.
- (1944): *Ethics and Language*. Yale University Press, New Haven.
- Strawson, P. (1950): "On referring", *Mind* 59, 320-344.
- (1952): *Introduction to logical theory*. Methuen, London.
- Stump, E. (1989): *Dialectic and its place in the development of medieval logic*. Cornell University Press, Ithaca.

- Vanderveken, D. (2002): "Universal Grammar and Speech act theory", in Vanderveken, D. & Kubo, S. (eds), *Essays in Speech Act Theory*. John Benjamins, Amsterdam, 25-62.
- Vendler, Z. (1964): "The Grammar of Goodness", *The Philosophical Review* 72 (4), 446-465.
- Walton, D. (1996): *Argumentation Schemes for Presumptive reasoning*. Lawrence Erlbaum, Mahwah.
- (2005): "Deceptive Arguments Containing Persuasive Language and Persuasive Definitions", *Argumentation* 19, 159-186.
- Walton, D. & Macagno, F. (2008): "Reasoning from Classifications and Definitions", *Argumentation* 23, 81-107.
- (2010): "Defeasible Classifications and Inferences from Definitions", *Informal Logic* 30, 34-61.
- Walton, D. & Krabbe, E. (1995): *Commitment in Dialogue*. State University of New York Press, Albany.
- Walton, D., Reed, C. & Macagno, F. (2008): *Argumentation Schemes*. Cambridge University Press, New York.
- Wilson, D. (1975): *Presupposition and Non-Truth-Conditional Semantics*. Academic Press, London.
- Wright, G. von (1972): "On so-called practical inference", *Acta Sociologica* 15, 39-53.
- (1963): *The varieties of goodness*. Routledge & K. Paul, London.
- Zarefsky, D. (1998): "Definitions", in Klumpp J. (ed.), *Argument in a time of change: Definitions, frameworks, and critiques*. National Communication Association, Annandale, 1-11.





# Présuppositions discursives, assertion d'arrière-plan et persuasion

Louis de SAUSSURE, Université de Neuchâtel

## 1. Introduction

On distingue traditionnellement entre: a) le contenu posé ou asserté, qui contient également les implications de la phrase, b) le contenu implicite, et c) le contenu présupposé. Le statut des présuppositions est largement problématique et donc débattu: à certaines conditions, les présuppositions n'émergent pas – elles ne «projettent» pas. En général, les présuppositions sont assimilées à l'*arrière-plan conversationnel*, dès lors qu'elles sont censées présenter des informations déjà connues des interlocuteurs, informations sur la base desquelles la communication pertinente est élaborée (Stalnaker 1974 pose le lien entre les présuppositions et l'*arrière-plan conversationnel*). Sperber & Wilson (1995: 217) suggèrent l'idée que la distinction entre *avant-plan / information nouvelle* et *arrière-plan / information ancienne* n'est pas intrinsèquement liée à la distinction entre contenus assertés et présupposés; au contraire, la qualité d'*arrière-plan* ou d'*avant-plan* résulte selon eux du traitement de l'énoncé. Ils ont raison en ceci que des informations nouvelles peuvent parfaitement être amenées par des présuppositions, et d'ailleurs cela arrive constamment. Ce phénomène est celui de l'*accommodation* présuppositionnelle, et ne pose aucun problème (voir à ce sujet Thomason 1990 et Thomason & al. 2006). Dans certains cas, il s'agit d'informations très peu pertinentes, mais qui servent indirectement à la pertinence générale de l'énoncé. Nous regarderons quelques exemples plus bas, mais si quelqu'un explique *Je ne peux pas venir à la réunion; je dois chercher mon chat chez le vétérinaire*, l'interlocuteur peut très bien ignorer l'existence du chat et l'apprendre incidemment par *accommodation* présuppositionnelle. Dans d'autres cas, l'*accommodation* présuppositionnelle constitue l'information véritablement pertinente, de manière d'ailleurs totalement ouverte, afin d'amener des effets

pragmatiques supplémentaires. Ainsi, lorsque Fidel Castro, du moins dans le film *Che*, nomme Guevara commandant, il le fait par une accommodation présuppositionnelle: «Signez ici, commandant», ce qui lui permet de suggérer que le statut de commandant va en quelque sorte de soi pour Guevara, qu'il s'agit d'un point hors de la discussion, etc., en exploitant ainsi les propriétés de la présupposition à des fins rhétoriques et pragmatiques. Un tel cas est d'ailleurs proche de ce que Karttunen (1974) appelle des *présuppositions informatives*, du type de *Nous regrettons de vous informer que les enfants ne peuvent accompagner leurs parents aux exercices* (notre traduction et adaptation), amenant à des actes de langage indirects.

Mais il existe un troisième type de cas où des présuppositions, et des accommodations de présuppositions, sont exploitées par un locuteur: il s'agit des situations où une information conflictuelle avec les connaissances du destinataire, voire une information absurde, est prise pour acquise dans la conversation sans même que le destinataire ne remarque qu'il vient d'incorporer une nouvelle information, peut-être absurde, dans son environnement cognitif. Dans ce cas, la présupposition est exploitée de manière «couverte», elle est communiquée de manière indirecte et selon des moyens qui signalent la nouvelle information comme ancienne.

Ainsi, la proposition de Sperber & Wilson de dissocier *arrière-plan* et *présupposition* ne doit pas être comprise comme faisant fi des propriétés des présuppositions; en utilisant une présupposition, un locuteur peut, les circonstances aidant, faire admettre une proposition comme relevant de l'arrière-plan. Les présuppositions incitent – mais n'obligent pas – à considérer l'information qu'elles portent comme étant déjà connue et partagée, et donc non pertinente dans l'échange conversationnel.

On remarque cependant que non seulement les présuppositions *stricto sensu*, qui sont déclenchées linguistiquement, entrent dans ce schéma, mais également des éléments de contenus qui sont purement inférés, que nous appellerons des *présuppositions discursives*. Elles sont en quelque sorte le pendant des implicatures conversationnelles du côté des présuppositions. Elles n'entrent pas dans ce qui constitue l'objectif communicatif du locuteur (ce qui les rapproche des présuppositions) et sont entièrement inférées (ce qui les assimile à des implicatures d'un certain type).

Lorsqu'une présupposition est amenée par une énonciation sans que son contenu n'appartienne en fait à l'arrière-plan conversationnel – à

L'environnement cognitif du destinataire, pour être plus précis –, cette présupposition peut se trouver rejetée, mais avec un fort coût interactionnel et cognitif, ou au contraire, et c'est le cas habituel, se trouver accommodée, c'est-à-dire se trouver intégrée à l'environnement cognitif. Néanmoins, dans ce cas, elle ne se trouve pas rejoindre l'arrière-plan de la même manière qu'une information posée ou implicite qui se présente comme constituant l'objectif de la communication verbale. Au contraire, une présupposition se présente comme située en dehors des éléments qui suscitent l'attente de pertinence, de sorte que l'attention ne s'y porte pas. Il se trouve que, à cause de cela, dans des contextes machiavéliens comme dans des contextes ordinaires, la présupposition peut se trouver accommodée sans évaluation critique; et même, elle peut se trouver accommodée tout en présentant des inconsistances majeures avec les autres éléments connus du destinataire. Dans ce dernier cas, l'accommodation présuppositionnelle constitue un outil efficace au sein de stratégies persuasives. C'est à étudier la relation entre pertinence et accommodation présuppositionnelle qu'est consacré cet article; il sera notamment suggéré que l'accommodation touche aussi des éléments de nature présuppositionnelle, mais qui ne sont en général pas considérés dans la littérature sur les présuppositions, et qui concernent les motivations implicites de l'énonciation d'un contenu propositionnel. Nous les appellerons *présuppositions discursives*, en développant l'hypothèse qu'elles sont particulièrement sensibles, car elles forment des prérequis pour la pertinence tout en échappant à l'évaluation en termes de pertinence. Nous établissons un lien entre leur comportement et une série de biais cognitifs en relation avec le caractère déjà connu des informations, suggérant que le fait qu'une information soit simplement présentée comme déjà connue – présupposée – suffit à la rendre efficace, même si elle n'est pas en réalité déjà connue. En somme, nous suggérons qu'une information *présentée comme ancienne* a de bonnes chances d'être effectivement traitée comme telle, même lorsque c'est indu, c'est-à-dire même lorsque l'information concernée est en fait nouvelle. Et même, lorsque cette information est inconsistante avec l'arrière-plan réel, l'inconsistance passant alors inaperçue, du fait que l'information ancienne n'a pas besoin d'être évaluée: son ancienneté l'assortit de la présomption qu'elle a déjà fait l'objet de toutes les évaluations utiles.

Dans un premier temps, nous proposerons de revoir la catégorisation habituelle des types de contenus pour y faire tenir ce que nous identifions ici

comme *présuppositions discursives* vis-à-vis des implications, des implicatures (ou «implications») et des présuppositions *stricto sensu*. Ensuite, nous aborderons la notion d'*accommodation présuppositionnelle* en suggérant qu'elle s'applique aux deux catégories de présuppositions que nous reconnaissons, avant de nous tourner vers deux illustrations d'accommodations indues, chacun jouant sur l'un des types de présuppositions retenus. Nous ferons au cours de notre démarche le lien entre l'accommodation présuppositionnelle et le traitement superficiel (*shallow processing*) de l'information, suggérant que l'accommodation présuppositionnelle repose sur un mécanisme de traitement superficiel de l'information, que nous relierons aux heuristiques et biais cognitifs de la tradition initiée par Tversky & Kahneman (1974). Greco (2003), discutant Walton (1999), suggère que la présupposition sert d'ancrage à des effets d'argumentation fallacieuse; reliant cette observation à la notion d'*illusion cognitive* (Pohl 2004), nous suggérerons l'existence de biais, ou *illusions*, ou encore *heuristiques* typiquement présuppositionnelles.

## 2. Présupposition discursive et types de contenus

La littérature reconnaît une variété de phénomènes présuppositionnels actifs au niveau du discours. Par exemple, les *conditions de félicité* austiniennes peuvent être considérées comme des sortes d'informations présupposées. Des présomptions au sujet des savoirs partagés, sans lesquels les énoncés ne peuvent prendre de sens concret, peuvent aussi être appelées *présuppositions* comme le propose Macagno (2012). Une notion de présupposition *lato sensu* est également posée par Rigotti & Rocci (2001), comme éléments qui contribuent à assurer la congruité de l'information dans le contexte discursif. Cette notion de présupposition sera considérée dans ses aspects persuasifs par Rigotti (2005), ainsi que par Rocci (2005); et van Dijk (1998) et (2011) appelle également *présuppositions* des prémisses implicites que nous appelons ici *présuppositions discursives*.

Toutefois, ces types d'informations ne peuvent pas être appelés *présuppositions* autrement que dans un sens non technique, car elles présentent des différences avec les présuppositions *stricto sensu* (notamment sur leur

insensibilité à la force illocutionnaire). Il n'en reste pas moins que l'on peut considérer que si ces informations sont implicites – ce sont des sortes d'implicatures –, elles sont également spéciales, en ce sens qu'elles ne se présentent pas comme l'objet de la communication, mais comme un savoir déjà présent. Nous allons ici tenter d'identifier plus clairement ces sortes d'informations en les associant à une notion de pertinence (dans la continuation de Simons 2005); pour nous, les présuppositions *lato sensu*, que nous appellerons *discursives*, ont le sens de pré-conditions, ou conditions nécessaires, non pas pour le sens véridictionnel (contrairement aux présuppositions *stricto sensu*) mais pour la pertinence de l'énoncé.

Nous allons rapidement en venir à quelques exemples, mais il convient d'abord de rappeler que la notion de *présupposition* souffre d'un certain flou dans la littérature dès que l'on considère les cas où elles émergent sans déclencheur linguistique clairement identifiable ni ne sont fondées sur une implicature conversationnelle (voir Gazdar 1979). Il s'agit alors de cas où une inférence pragmatique est nécessaire comme condition pour que l'énoncé puisse avoir une quelconque pertinence, et cette inférence partage des propriétés cruciales avec les présuppositions *stricto sensu* plutôt qu'avec les implicatures conversationnelles. De telles inférences sont des prérequis sans lesquels la phrase conserve certes un sens véridictionnel, mais ne peut pas avoir la moindre pertinence. En même temps, ces informations ne sont pas présentées par le locuteur comme formant l'objet de la communication. Sans elles, il y aurait un sens, mais l'acte de langage serait absurde. Elles sont à propos d'informations réputées préalables et donc ne sont pas des implicatures conversationnelles au sens classique, c'est-à-dire qu'elles ne sont pas ce que le locuteur manifeste comme son vouloir-dire.

Interrogeons-nous par exemple à propos du statut des inférences (1a-1b) sur la base de (1), ou (2a-2c) sur la base de (2):

- (1) Il faut fermer la porte / Ferme la porte!
- (1a) L'ouverture de la porte est non désirable dans les circonstances
- (1b) Il y a trop de bruit / Il fait froid / Nous manquons d'intimité...
- (2) Les armes sont interdites dans ce secteur.
- (2a) Les armes auraient pu être autorisées dans ce secteur.
- (2b) Les armes peuvent être autorisées dans d'autres secteurs
- (2c) Les armes sont indésirables / dangereuses si portées dans ce secteur

Les actes de langage (1) et (2) reposent sur un certain nombre de pré-conditions qui expliquent ou justifient leur occurrence en discours. (1a) n'est aucunement une présupposition dans le sens usuel, car ce contenu n'est pas inféré sur la base d'un item lexical ni même de la signification de la phrase, et n'entre pas même en ligne de compte pour le calcul du sens: les conditions du monde, pas plus que les états mentaux des interlocuteurs, ne modifient la signification des phrases, hormis pour ce qui est de la référence et de la désambiguïsation. Que le caractère ouvert de la porte soit désirable ou indésirable n'affecte en rien le sens de (1). Des inférences comme celles mentionnées ci-dessus sont dérivées sur la base de connaissances encyclopédiques et de compétences pragmatiques – notamment la capacité d'attribuer au locuteur des intentions de sens pertinentes dans le contexte – au sujet de ce qui justifie une requête comme (1) ou une injonction comme (2). On relèvera que des inférences comme (1b) (il y a du bruit, etc.) sont d'un niveau d'accessibilité plus éloigné que (1a), car elles forment les raisons justifiant (1a), qui est lui-même un arrière-plan justifiant (1). De même, (2) n'a nullement besoin de convoquer (2a-c) pour avoir une signification claire et complète. De plus, (2) ne sert normalement pas à communiquer des propositions comme (2a-c), mais tout simplement à communiquer l'interdiction des armes au lieu concerné. Evidemment que de telles propositions sont bien des implicatures d'une certaine sorte (ce sont des inférences complètement pragmatiques), tout comme, pense-t-on souvent, les présuppositions *stricto sensu*; cependant, il ne s'agit ni de présuppositions *stricto sensu* (indispensables à la signification) ni d'implicatures conversationnelles (identifiées comme formant l'objet, ou un objet, de la communication intentionnelle).

Plutôt que de les penser comme pré-conditions pour des actes de langage, nous proposons de les voir comme des inférences nécessaires pour faire sens de l'intention informative du locuteur, qui survient dans un certain monde, doté de certaines propriétés ontologiques (ou au moins dans un monde ainsi représenté dans l'esprit des interlocuteurs). Un panneau indiquant (2), s'il était installé à l'Université de Neuchâtel, signifierait toujours la même chose (les armes sont interdites en ce lieu), et certainement que l'acte de langage serait efficace (il assurerait l'interdiction), même s'il ne serait pas *heureux* au sens searlien qui exclut que le contenu de l'acte de langage soit évident pour le destinataire. Cependant, il serait absurde, du fait que les armes sont de

toute manière interdites sur la voie publique dans ce pays, et le panneau serait interprété comme un gag.

Des éléments comme (1a-b) et (2a-c) sont des bases préliminaires sans lesquelles (1) et (2) ne sauraient prétendre satisfaire les attentes du destinataire en termes de rationalité de la communication, ou de présomption de pertinence (Sperber & Wilson 1995).<sup>1</sup> Nous appellerons de telles propositions des *présuppositions discursives*. Elles pourraient peut-être s'assimiler à des prémisses contextuelles soutenant l'acte d'assertion principale, lui donnant ses raisons de survenir dans le monde.

Dans les situations les plus courantes, de telles informations sont connues du destinataire dès avant l'acte de parole, et donc appartiennent à l'arrière-plan conversationnel. Le locuteur anticipe l'existence d'un arrière-plan partagé (*common ground*) sur lequel la conversation s'élabore (Clark & al. 1983); un tel partage d'informations préalables est même une condition pour la capacité que nous avons de prédire le sens que nos interlocuteurs seront capables d'extraire de nos énoncés, et conditionne la valeur ajoutée de toute nouvelle information dans le discours. Un *common ground* est une base nécessaire qui évite aux interlocuteurs de devoir constamment rétablir les fondements sur lesquels leurs pensées s'élaborent et se communiquent; il n'est donc pas surprenant que la distinction entre avant- et arrière-plan soit considérée comme pertinente sur le plan cognitif (au moins depuis Haviland & Clark 1974). Les présuppositions discursives sont des éléments qui appartiennent aux connaissances d'arrière-plan, mais qui sont nécessaires non pour le sens mais pour la pertinence. Sperber & Wilson présentent une conception hétérodoxe pour la sémantique formelle en suggérant que la distinction entre *asserté* et *présupposé* ne recouvre pas celle qui sépare l'*arrière-plan* de l'*avant-plan*. Pour eux, ces dernières notions ne sont aucunement linguistiques; c'est la recherche de pertinence qui amène un individu à positionner une information comme relevant de l'*avant-* ou de l'*arrière-plan*, quelle que soit la manière dont cette information est amenée dans l'énoncé. L'information d'arrière-plan, expliquent-ils, est une information «qui ne contribue à la pertinence que de manière indirecte, en réduisant l'effort de traitement requis» (Sperber & Wilson 1995: 217, notre traduction). Cela

1 Simons (2005) suggère que les présuppositions, tant au sens classique que les hypothèses contextuelles implicites, servent à établir la pertinence.

n'implique pas qu'il serait impossible pour le locuteur d'utiliser des moyens linguistiques pour communiquer qu'il considère une information comme relevant de l'arrière-plan (la sémantique des déclencheurs linguistiques de présuppositions y invite), ou des moyens pragmatiques, quand il peut anticiper le fait que certaines propositions nécessaires pour faire sens de l'énoncé doivent être tenues pour vraies préalablement par tout interlocuteur dans les circonstances. C'est, suggérons-nous, le rôle des présuppositions discursives. Mais le fait que les présuppositions ne sont pas *toujours* de l'arrière-plan complique encore les choses.

Des éléments de l'arrière-plan conversationnel sont généralement récupérés par des moyens présuppositionnels en vue de produire le contexte dans lequel la nouvelle information sera traitée. Du fait que sous certaines conditions les présuppositions sont bloquées (elles ne «projettent» pas), des chercheurs comme Carston (1998) considèrent qu'elles sont de nature pragmatique et donc générées en contexte plutôt que produites par défaut puis éventuellement annulées sous des conditions spécifiques. Ainsi, il existe un ancien débat, qui n'est toujours pas définitivement clos, pour savoir si les présuppositions, même en présence d'un déclencheur linguistique, sont plutôt sémantiques (déclenchées par défaut par l'expression) ou pragmatiques (étant plutôt le résultat d'un processus inférentiel qui les réalise ou ne les réalise pas en fonction de principes pragmatiques). Une des raisons fondamentales de cette hésitation de la littérature, outre que les présuppositions, parfois, ne «projettent» pas, réside dans le fait qu'elles sont généralement en lien direct avec un élément sémantique codé (le «déclencheur»), mais qu'en même temps elles émergent aussi pour la résolution des anaphores, laquelle est également sujet à l'accommodation pragmatique, et que les théories sémantiques et pragmatiques regardent différemment (voir Heim 1983, van der Sandt 1992, Karttunen 1974, mais la littérature sur cette question est énorme).

Trois grands types de contenus sont généralement distingués: a) le contenu asserté (ou *posé*), qui est assimilables aux *implications*; b) le contenu implicite, les *implicatures* ou *implications* de la tradition gricéenne; et c) le contenu *présupposé*. La théorie de la Pertinence de Sperber & Wilson (1986/1995) présente un tableau différent, pour lequel l'opposition pertinente ne se situe pas entre un contenu «dit», vériconditionnel, et un contenu «implicité» et non-vériconditionnel, mais entre des niveaux explicites



et implicites de communication qui sont récupérés pragmatiquement sur la base d'une forme abstraite et de principes pragmatiques qui conduisent à former une hypothèse sur l'intention informative du locuteur. Considérées par la lorgnette de la théorie de la Pertinence, ce que nous appelons des *présuppositions discursives* n'est finalement rien d'autre que des implicatures d'un genre spécifique qui se présentent elles-mêmes comme *a priori*, comme de l'information préalable nécessairement incorporée à l'interprétation pour pouvoir donner un sens minimalement pertinent à l'acte d'un locuteur de présenter une certaine information dans les circonstances considérées.

De ce point de vue, nous obtenons:

- (a) des éléments assertés / impliqués, qui sont manifestés comme faisant ouvertement partie de la communication pertinente et sont déterminés par les relations logiques des items linguistiques (comme *tuer* qui implique *être mort* pour le patient);
- (b) des implicatures, qui sont manifestés comme faisant ouvertement partie de la communication pertinente et sont déterminés pragmatiquement;
- (c) des présuppositions *stricto sensu* qui ne sont pas manifestées comme ouvertement pertinentes et sont déclenchées linguistiquement;<sup>2</sup>
- (d) des présuppositions *discursives*, qui ne sont pas manifestées comme ouvertement pertinentes et sont déclenchées pragmatiquement.<sup>3</sup>

Dans ce qui suit, nous tenterons d'étendre la notion d'*accommodation présuppositionnelle* aux présuppositions discursives avant d'observer deux cas d'exploitation manipulateur de présuppositions.

2 Nous admettons ici que les présuppositions qui dérivent pragmatiquement d'une expression linguistique (*quasi-présuppositions* chez Schlenker 2008) sont de ce type. Par exemple, *Jean n'est pas venu en retard* présuppose en quelque sorte faiblement que Jean est venu, sur la base du déclencheur linguistique *en retard* qui est focalisé (voir aussi Abrusan 2013).

3 Dans son excellent article sur la question de la présupposition, Kleiber (2012) distingue entre *présuppositions contenus* et *présuppositions inférences*. Pour lui, il s'agit de différents niveaux reliés; nous parlons ici *a priori* d'éléments distincts.

### 3. Remarques sur l'accommodation présuppositionnelle

Les présuppositions posent trois types majeurs de problèmes quant à leur survenue dans le discours.

D'abord, elles sont tirées comme des informations de l'arrière-plan connu, mais il se peut très bien, en fait, qu'elles ne soient pas telles, auquel cas elles donnent lieu à une accommodation, c'est-à-dire d'inclusion dans l'environnement cognitif du destinataire (ou de «mise à jour du contexte»). Nous postulons qu'un phénomène d'accommodation survient également avec les présuppositions discursives; cependant, il convient de comprendre la nature de ce processus inférentiel qui n'est ni celui de la présupposition *stricto-sensu*, qui repose sur des éléments linguistiques (ou des implicatures, cf. Gazdar 1979) ni celui des implicatures, puisque ces dernières relèvent de l'objectif communicationnel manifesté par le locuteur.

Deuxièmement, dans un tel cas, il est difficile de dire qu'elles ne sont pas pertinentes: elles amènent une information nouvelle, et elles le font d'une manière d'autant plus économique qu'elles épargnent à la cognition le recours à un processing profond d'évaluation critique. De ce point de vue, de telles présuppositions, accommodées, sont hautement pertinentes: des effets cognitifs (informationnels) potentiellement forts accompagnent un coût de traitement minimal. Pourtant elles n'entrent manifestement pas dans les attentes de pertinence, qui sont focalisées sur de l'information réputée nouvelle et non réputée ancienne. Troisièmement, enfin, leur accommodation peut se produire de manière inconsistante, donc erronée du point de vue de l'environnement cognitif. Même des présuppositions parfaitement fausses de manière en principe évidente peuvent ne pas être identifiées comme telles (nous verrons deux cas plus précisément plus bas). Dans les situations de communication ordinaire, la cognition réagit aux présuppositions jusqu'à un certain point seulement (Lewis 1979).

Pour aborder ces difficultés, en particulier celles qui sont posées par les présuppositions discursives, il faut commencer par reprendre les propriétés formelles des présuppositions. Elles sont discutées dans la littérature au moins depuis Frege (sans revenir à Thomas d'Aquin ou aux arguments éristiques d'Aristote) et, en ce qui concerne la portée de la négation, bien entendu, depuis les travaux de Russell ou Strawson. Du côté du discours, la

notion cruciale reste celle d'accommodation, produite par Lewis (1979), et, de manière un peu différente, par Ducrot (1972: 51) sous le terme de *coup de force présuppositionnel*, qui consiste à faire passer comme argument un contenu présupposé. L'idée de Lewis est, soit dit à gros traits, que l'accommodation présuppositionnelle (qui consiste en l'incorporation dans l'arrière-plan d'une présupposition qui n'y figurait en réalité pas) se produit mais «dans certaines limites» (Lewis 1979: 417, notre trad.). Le cas typique de l'accommodation présuppositionnelle est celle où une résolution anaphorique appelle l'introduction d'un référent qui n'a pas été mentionné préalablement. Dans divers autres cas de figure, l'introduction d'une présupposition n'est pas nécessaire pour obtenir un sens en soi (une signification propositionnelle saturée) mais pour obtenir un sens pertinent: c'est le cas de nos présuppositions discursives. Elles donnent les raisons contextuelles de l'énonciation, qui peuvent à leur tour avoir un impact sur les contenus implicites (implicatures) ou, au niveau discursif, sur diverses croyances associées à la compréhension de l'énoncé. L'extension du domaine de la présupposition au niveau discursif avait déjà été sommairement proposée par van Dijk (1998) avec le dialogue suivant, sur lequel Rocci (2005) élabore l'idée cruciale selon laquelle de tels contenus présupposés ont à voir avec une notion de cohérence ou de «congruité»:

- (3) A- Un voleur s'est introduit dans la maison de tante Julie.  
 B- Ce n'est pas surprenant. N'as-tu pas vu tous ces gens de couleur qui traînent par ici ces temps-ci? (van Dijk 1998: 206, notre traduction).
- (4) Présupposition: *Les gens de couleur sont des voleurs.*

Nous proposons donc d'apporter l'idée que (4) est produit aux fins de pertinence. L'implicature générée par B a à voir avec la politique, avec l'idée qu'il est souhaitable de renvoyer les gens de couleur chez eux, etc., mais le contenu (4) est indispensable à la contextualisation de (3B). Ce type de prémisse implicite, enthymématique, est de l'ordre de la présupposition, car elle est introduite dans le discours comme fait déjà manifeste, *tenue pour vraie* préalablement à l'énonciation.

Certaines accommodations présuppositionnelles sont «pragmatiquement inappropriées»; von Stechow (2000 et 2008) discute à cet égard la paire d'exemples suivante:

- (5) Je ne peux pas venir à la réunion: je dois chercher mon chat chez le vétérinaire.
- (5') Présupposition: *Le locuteur possède un chat.*
- (6) Oh, papa, j'ai oublié de te dire que mon fiancé et moi allons déménager à Seattle la semaine prochaine (von Fintel 2000: 15, notre trad.)
- (6') Présupposition: *La locutrice a un fiancé.*

La différence critique entre les exemples (5) et (6), dans le cas où l'information présupposée n'est *pas* déjà dans l'arrière-plan conversationnel, est que (5') est pragmatiquement «appropriée» tandis que (6') ne l'est pas – en supposant un contexte neutre ou ordinaire où le fait d'apprendre que le locuteur a un chat n'a pas de conséquences particulières (donc n'est pas pertinent en soi) tandis que d'apprendre l'existence d'un fiancé de sa fille en a. Les diverses théories qui traitent de la notion d'arrière-plan partagé (*common ground*) suggèrent toutes qu'il y a des principes généraux présidant à l'acceptabilité des présuppositions. Une manière fine de poser les choses est donnée par von Fintel (2000: 15): «La théorie du *common ground* [...] dit que quelqu'un peut [*can*: «peut» dans un sens radical ou déontique] présupposer P dès lors qu'il est assumé que le destinataire ne va pas vouloir contrer [*dispute*] P» (notre trad.). Pour les besoins de la discussion que nous tenons ici, nous pouvons nous en tenir à cette idée d'un savoir partagé ou mutuel; néanmoins, s'il fallait être plus précis, il faudrait comprendre que le locuteur ne peut amener une présupposition menant à une accommodation que s'il fait l'hypothèse que l'interlocuteur n'aura pas de raison de discuter la vérité de P. Dans le cas contraire, la communication fait intervenir une accommodation présuppositionnelle inappropriée.

Cependant, tout ceci laisse de côté le fait que les présuppositions peuvent se comporter différemment selon l'environnement cognitif du destinataire. Ce fait affecte directement le caractère approprié des accommodations présuppositionnelles. Il est en réalité nécessaire de prendre en compte deux conditions nécessaires pour le caractère pragmatiquement inapproprié d'accommodations présuppositionnelles du type de (6).

Premièrement, le fait que (6') contredit une hypothèse contextuelle du destinataire (*la locutrice n'a pas de fiancé*) doit être comparé au fait que (5') introduit une information qui ne présente aucune inconsistance avec un savoir préalable: qu'un employé ait ou non un chat ne figure pas comme hypothèse manifeste à l'interlocuteur de (5); l'interlocuteur est «agnostique»

en la matière (von Stechow 2000: 11). De la sorte, l'existence du chat est donc indifférente, car dépourvue de conséquence pratique (n'importe quelle autre raison de ne pas venir à la réunion aurait eu exactement les mêmes conséquences pratiques dans l'échange conversationnel), ce qui conduit donc à l'absence de raison de contrer (5'), par comparaison avec (6').

Deuxièmement, mais de manière immédiatement reliée au point précédent, il est, dans des circonstances dont nous parlons, dépourvu de pertinence que l'employé ait un chat, par comparaison avec n'importe quelle autre raison valable de ne pas venir à la réunion, mais il est très pertinent en soi que la locutrice ait un fiancé.

La pertinence, pour un individu, réside dans l'impact produit par une nouvelle information dans sa connexion avec d'autres informations préalablement entretenues, de sorte que de nouvelles informations sont déduites, et produisent de l'effet; ceci, à la mesure de l'effort dépensé pour extraire ces informations (Sperber & Wilson 1986; 1995).

Dans les exemples ci-dessus, même s'il y a une incorporation dans l'environnement cognitif du destinataire avant un quelconque traitement critique, la possibilité de réagir existe. Mais il y a un nombre de cas où cette possibilité est sérieusement obérée, et une inconsistance quelconque passe alors inaperçue. Il suffit pour cela de s'éloigner de la seule présupposition existentielle pour aborder des cas moins évidents.

Or nous verrons plus bas que dans diverses situations de communication, l'accommodation de la présupposition présente des anomalies majeures qui ne sont pas détectées, ouvrant une vulnérabilité cognitive sérieuse à des effets de persuasion et de manipulation. Les théories de la présupposition et du *common ground* laissent ce problème de côté, avec raison d'ailleurs, puisqu'elles portent sur des faits de sens, tandis que la question que nous soulevons concerne le traitement général de l'information par la cognition.

En fait, c'est bien connu, quand un rejet présuppositionnel intervient, c'est toujours au prix d'un effort significatif. Cet effort est directement dû au fait que la présupposition se présente comme une information ancienne: d'une part, elle ne s'assortit pas d'un traitement en termes de pertinence, puisqu'elle est censée préexister d'une manière ou d'une autre et donc avoir déjà subi un traitement, et, d'autre part, du fait même qu'elle est présentée comme déjà connue, elle est présentée comme partagée, ou mutuellement manifeste, au locuteur et à l'interlocuteur (ou, cela va sans dire, aux

interlocuteurs); il s'ensuit que contredire une présupposition non seulement demande de se focaliser sur ce qui n'est pas focalisé dans l'énoncé, mais implique de remettre en question les bases communes de la conversation, le savoir réputé mutuel, déjà acquis, ce qui présente un inconvénient de *face*, pour le dire en termes goffmaniens, ou, plus simplement, risque de faire passer l'interlocuteur pour un ignorant ou un individu qui se démarque du groupe fondé sur un savoir mutuellement partagé.

Si l'on admet que le traitement peut être *profond* – pour ce qui concerne les contenus posés – c'est qu'il peut être *superficiel* aussi. On aura compris que la position qu'on défendra ici implique que le traitement des présuppositions est un traitement superficiel. La littérature (cf. plus bas) décrit le traitement de certaines informations linguistiques posant des problèmes de consistances par le terme de *shallow processing*, ou «traitement superficiel», typiquement lorsqu'un item lexical est réinterprété comme concept de la classe d'ordre supérieure. Il est possible que l'accommodation présuppositionnelle soit elle aussi sujette à l'accommodation pragmatique superficielle, mais nous verrons plus bas que cette hypothèse n'explique de loin pas tout. Avant cela, il convient de détailler les raisons qui nous conduisent à suggérer que les présuppositions de manière générale échappent aux contrôles de pertinence.

#### 4. Présuppositions et pertinence

Les présuppositions ont des propriétés sémantiques; les plus évidentes concernent leur rapport à la négation, puisqu'elles sont insensibles à la polarité de l'énoncé (ainsi qu'à la force illocutionnaire en général). Mais les présuppositions ont également des propriétés discursives connues: elles tendent à échapper aux mouvements discursifs (c'est-à-dire qu'elles sont normalement laissées sans réaction). Les propriétés sémantiques et discursives des présuppositions vont main dans la main, de sorte que leur comportement discursif, comme leurs propriétés sémantiques, sont le produit de quelque chose de plus profond, et dont nous avons déjà parlé, à savoir que les présuppositions se présentent comme des faits déjà établis. Les présuppositions sont vraies indépendamment de la valeur de vérité du

contenu propositionnel lui-même, et donc leur valeur de vérité reste en dehors de tout questionnement ou évaluation. On sait depuis Ducrot (1972) que les présuppositions sont en effet non véridicibles, de sorte qu'une réponse comme *c'est faux* ne peut porter sur une présupposition. Cela apparaît particulièrement clairement avec les questions, seul le contenu faisant l'objet de la question pouvant donner lieu à un enchaînement discursif régulier:

- (7) A- Est-ce que Jean a raté son examen?  
 B- \*C'est faux, il est resté à la maison.  
 B- Oui / Non.

Il y a bien des ressources langagières autorisant le changement de focalisation du contenu posé vers le contenu présupposé, mais elles sont indirectes, comme avec *Attends une minute: il ne s'est pas présenté*. Ce type d'énoncé peut s'analyser comme cassant une norme admise de la communication, ne respectant pas un schéma de question-réponse, soulevant un problème de face complexe où le locuteur est montré comme ayant commis l'erreur de prendre pour acquis ce qui ne devait pas l'être.

Néanmoins, une explication beaucoup plus terre-à-terre est aussi disponible, éventuellement en complément de celle-ci: défaire une présupposition ou la questionner revient à mettre en doute la pertinence même de l'énoncé qui la portait, voire à la nier.

Comme la présupposition doit être vraie pour que l'énoncé puisse avoir un sens complet, mettre en doute la présupposition revient à mettre en doute le sens et donc la pertinence de la question. Au niveau des présuppositions discursives, auxquelles nous allons revenir sous peu, les mettre en doute revient à mettre en doute les motivations factuelles conduisant à l'énonciation qui les requiert; de la sorte, l'énoncé se trouve dénoncé comme privé de pertinence. Il y a donc là un coût particulier lié à une communication qui devient soudain métalinguistique ou métadiscursive, et qui pourrait ressembler à certains égards à l'*opting out* de Grice. Il n'y a donc rien de surprenant à ce que ce soit un mouvement défavorisé dans l'échange discursif.

Ducrot (1972) a également montré que les présuppositions sont rétives à une connexion par une conjonction, à l'exception de *et* et *si*:

- (8) Jean a cessé de fumer...  
 a. ...parce qu'il souhaite être en bonne santé.  
 b. ...\*parce qu'il aimait ça.

(8b) est impossible – hormis dans une lecture très différente, parenthétique et métalinguistique – bien qu’il y ait une connexion conceptuelle immédiatement disponible entre le fait de fumer et d’aimer ça. Mais comme l’assertion est à propos de cesser de fumer, qui ne fait que présupposer *fumer*, l’enchaînement n’est plus possible, *parce que* ne pouvant enchaîner que sur le contenu propositionnel posé. L’intéressant est ailleurs: (8b) n’est pas seulement inattendu ou étrange en conversation, il est tout à fait *impossible* en dehors de la lecture parenthétique. Nous prenons ceci comme une indication de la rigidité du confinement des présuppositions hors du domaine véridicible de la communication linguistique.

Une explication classique de ce comportement est que les présuppositions ne sont pas que sémantiques, elles sont aussi pragmatiques, puisqu’elles ont à voir avec le caractère ancien de l’information, l’arrière-plan, en conséquence de quoi elles ne sont pas sujettes à discussion (cf. aussi von Stechow 2008: 138).

En ce qui concerne l’accommodation des présuppositions, maintenant, elle est parfois comparée à l’implicature en tant que dispositif économique de la communication lié à l’inférence (Seuren 2000: 180, Stalnaker 2002: 105). Néanmoins, il y a des différences frappantes entre l’implicature et l’accommodation présuppositionnelle. En particulier, bien que les deux amènent une information supplémentaire, les implicatures le font de manière ouverte, de sorte qu’elles sont appréhendées comme des composantes de la communication intentionnelle, à laquelle la pertinence est attribuée, et même comme les composantes les plus pertinentes de la communication intentionnelle. À l’inverse, l’accommodation des présuppositions survient de manière couverte, puisque la présupposition est réputée non pertinente, ancienne, hors de propos. Néanmoins, les présuppositions, même amenées par accommodation, entrent sans difficulté au titre de prémisses dans des schémas argumentatifs (tout comme, bien évidemment, les contenus explicites et implicites). En (9) ci-dessous, l’assertion b) est justifiée par la proposition c) elle-même introduite par une possible accommodation présuppositionnelle (si le destinataire n’est pas au courant du tabagisme du locuteur avant le présent de l’énonciation):<sup>4</sup>

- (9) a) Jean a cessé de fumer. b) Il toussait chaque matin.  
c) *Jean fumait.*

4 Ou: la proposition b) est expliquée par la proposition c)...



Nous dirons que l'accommodation présuppositionnelle exige l'adhésion du destinataire sans exiger son consentement. L'accommodation présuppositionnelle, en ce sens, ignore les contrôles de pertinence. Une courte élaboration sur ce point est toutefois nécessaire.

Attribuer de la pertinence à un énoncé est normalement le résultat d'un processus assez simple (Sperber & Wilson 1986, 1995): l'information est inférée de l'énoncé de telle sorte qu'elle compense optimalement l'effort de traitement. Il y a donc un contrôle qui s'opère et s'applique à la fois aux *effets* du traitement et à son *coût*, et qui établit le moment où l'effet atteint le seuil d'équilibre entre les deux variables à partir duquel les attentes de pertinence sont satisfaites,<sup>5</sup> moment auquel le destinataire établit la plausibilité de son interprétation. Les présuppositions, étendues comme nous le proposons aux présuppositions discursives, c'est-à-dire aux motivations factuelles de production d'un énoncé dans les circonstances, forment le fondement implicite sur lequel un acte de langage est produit. Les présuppositions sont donc actives en situant le seuil de pertinence plus haut que zéro, pour ainsi dire, mais ne sont pas appréhendées comme formant le point en jeu dans la communication. Or, ceci se produit aussi avec des présuppositions indues qui se trouvent accommodées pour préserver, dans l'esprit du destinataire, la présomption que l'énoncé est pertinent. Car, sans de telles présuppositions, la pertinence peut dans certains cas disparaître complètement; or la présomption de pertinence est très forte, puisqu'elle conduit même occasionnellement à une accommodation conceptuelle automatique, sans que celle-ci soit même remarquée, lors de traitements superficiels d'informations conceptuelles (nous y revenons plus bas).

- 5 Par *effort*, on entend, pour la genèse des informations non explicites, la difficulté de la contextualisation: l'énoncé devant entrer dans un raisonnement conjointement avec d'autres prémisses, un nombre plus élevé de ces dernières équivaut à un effort plus grand. Pour ce qui est du niveau explicite, le nombre de données de contextualisation nécessaires aux opérations de précision du sens explicite est pertinent. Il s'agit en réalité d'un mécanisme qui se développe dans le temps, fût-il très court, du processus de compréhension, l'effort étant consenti avec une *anticipation* d'effet, la compréhension menant à des inférences supplémentaires tant que l'anticipation, qui augmente avec les éventuels surcroûts d'efforts, n'est pas satisfaite. Ce modèle fait l'objet de nombreux travaux expérimentaux en pragmatique cognitive et neuropsycholinguistique expérimentale.

Que les présuppositions discursives puissent être traitées à la manière des autres présuppositions en ce qui concerne leurs propriétés fondamentales ne va pas de soi. Notons premièrement que, selon les contextes, ce que nous avons appelé *présupposition discursive* peut très bien correspondre dans d'autres situations à des implicatures parfaitement standard. Par exemple, un énoncé comme (10):

(10) J'ai pris le parapluie.

peut soit présupposer (11) discursivement, dans l'idée que (11) forme une motivation factuelle assez naturelle pour (10), soit l'impliciter conversationnellement, dans la situation où le locuteur de (10) chercherait par son énoncé à rendre (11) manifeste:

(11) Il pleut / il est plausible qu'il pleuve aujourd'hui.

(11) est donc soit une présupposition discursive soit une implicature de (10), selon que l'échange conversationnel porte sur le parapluie ou sur la pluie.

Dans le cas présuppositionnel, (10) ne fait que s'élaborer à partir de l'hypothèse que (11) est déjà manifeste à l'interlocuteur, et dans ce cas, (11) n'est pas interprété comme satisfaisant les attentes de pertinence. Qu'un contenu comme (11) soit interprété comme implicature ou présupposition discursive dépend entièrement de paramètres pragmatiques: des tests formels comme celui de l'insensibilité de la présupposition à la négation ne montreront pas de stabilité sauf si nous savons déjà que l'information est à prendre comme implicature ou présupposition, ce qui n'est donc plus probant ou perd son objet.

Néanmoins, si on a affaire à une présupposition et non une implicature, on pourra constater le fait que (10<sup>\*)</sup> (la négation de (10)):

(10<sup>\*)</sup> Je n'ai pas pris le parapluie.

est également compatible avec les contextes où (11) est connu dans l'arrière-plan (la pluie tombe ou est attendue). Ainsi, (10) et (10<sup>\*)</sup> sont tous deux compatibles avec (11), de même que toute présupposition P de Q est compatible avec non-Q, tandis que si (10) sert à communiquer (11) comme implicature, (10<sup>\*)</sup> ne le peut pas.<sup>6</sup>

6 Sauf contexte rocambolesque où il est notoire que le locuteur aime se faire tremper en marchant.

Au niveau discursif, ceci correspond, suggérons-nous, à l'insensibilité sémantique de la présupposition à la négation, si nous admettons qu'au niveau explicite, la valeur de vérité de la présupposition n'est pas affectée par la valeur positive ou négative de la proposition *du fait qu'il s'agit d'un préalable*. Au niveau des présuppositions discursives, le fait qu'il s'agit d'un préalable entièrement pragmatique conduit à la compatibilité de contextes opposés avec l'énoncé. Autrement dit, les présuppositions discursives croisent une propriété des présuppositions – la compatibilité avec des assertions opposées – et une propriété des implicatures (conversationnelles gricéennes), le caractère optionnel et annulable.

Bien entendu, l'exemple choisi est illustratif. Dans le cas d'un énoncé du type *J'ai pris la voiture*, une quantité de présuppositions discursives peuvent servir de prémisses implicites pour la proposition explicitée (le locuteur est fatigué, la destination est lointaine, la valise est lourde...). La sélection de la bonne prémisses présupposée relève de processus qu'il est impossible d'esquisser pour l'instant, mais repose sur des aptitudes pragmatiques très générales qui concernent la contextualisation des énoncés et la détection de l'objectif communicatif.<sup>7</sup>

Nos présuppositions discursives, activées pour suggérer la motivation pragmatique de l'acte de langage, rendues disponibles pour que l'acte de langage puisse être pertinent, ont donc des propriétés fondamentalement communes avec les présuppositions *stricto sensu* comme les présuppositions d'existence, par exemple du chat en (5) ou du fiancé en (6).

Cela n'empêche pas les présuppositions discursives d'être éloignées du code à la manière des implicatures. Elles ne sont aucunement obligatoires pour la production du sens (contrairement aux présuppositions sémantiques et aux explicatures), tout comme les implicatures; elles sont donc annulables sans recourir à de complexes dispositifs comme la portée large de la négation. En même temps, les présuppositions discursives se comportent comme d'autres types de présuppositions, en ce sens qu'elles s'activent à un degré moindre de conscience, échappent à la focalisation et donc sont assorties en

7 Un relecteur se demande si (11) doit se voir comme cause du fait (10), ce qui relèverait de la «cohérence», ou comme raison de son énonciation. Cette distinction n'est pas opératoire ici: tout ce qui importe, c'est que (11) serve de prémisses implicite pour (10), ce qui signifie que (11) est un fait que le destinataire est amené à se représenter pour donner de la pertinence à (10).

quelque sorte d'une présomption de non-pertinence au profit du contenu posé en propre.

Et paradoxalement, en même temps, leur accommodation est en fait parfaitement pertinente, du fait même qu'elles produisent un effet cognitif sans passer par les contrôles évaluatifs coûteux en effort. Le paradoxe n'est en fait qu'apparent: lorsqu'elles sont accommodées, elles ne sont pas *vérifiées* dans leur pertinence, mais elles sont pertinentes. En revanche, lorsqu'elles ne sont pas accommodées, elles ne sont en effet pas pertinentes de leur propre chef, car leur évaluation donnerait un résultat déjà connu. Il n'est donc pas surprenant que le système tende à ne pas les vérifier.

Mais comme la pertinence inclut divers éléments comme la consistance, une information qui échappe aux contrôles de pertinence peut être inconsistante, et de manière plus globale, échappe aux contrôles critiques. Nous suggérons que c'est précisément cela qui forme le fondement causal de leur pouvoir persuasif dans les situations d'accommodation induite.

Nous avons suggéré plus haut que le fait que de telles accommodations puissent se produire de manière inaperçue peut avoir quelque chose à voir avec les effets, connus, de traitement *superficiel* de l'information. Nous discutons cette question ci-dessous à la lumière de deux exemples de tels actes de langage.

## 5. Pertinence d'arrière-plan: deux exemples

Lorsqu'une information présupposée est pertinente, mais que sa forme présupposée fait échapper sa pertinence au destinataire, qui la range dans la catégorie de l'arrière-plan sans autre forme de procès, nous parlerons d'*assertion d'arrière-plan*. Dans ce qui suit, nous observons deux cas d'assertion d'arrière-plan qui conduisent à l'incorporation dans l'environnement cognitif de croyances conflictuelles vis-à-vis d'autres éléments solidement établis. Le premier cas porte sur une présupposition *stricto sensu* et le second fait intervenir des présuppositions discursives.

Oswald (2010) rapporte un cas de figure saisissant.<sup>8</sup> Le journaliste français Philippe Vandel, humoriste à ses heures, a piégé le public en 1999 de façon spectaculaire. Lors d'un micro-trottoir, avec une caméra, il aborde les passants, leur explique que les spécialistes ont calculé que le Nouvel An 2000 aurait lieu un vendredi 13, et leur demande si cela les effraie. Les personnes interrogées et montrées dans le programme répondent à la question: elles expliquent qu'elles sont, ou non, effrayées, mais elles ne réagissent pas à l'information elle-même, qui pourtant porte une inconsistance majeure et qui devrait être évidente: un 1<sup>er</sup> de l'an ne peut être un 13.<sup>9</sup> Oswald (2010) suggère que l'expression *vendredi 13* est traitée de manière superficielle (ou «furtive», *shallow processed*) comme «un jour néfaste, superstitieux», la date elle-même n'étant pas traitée sur le plan calendaire. Il est remarquable d'ailleurs qu'à certains, Vandel ne se contente pas de dire que le «Nouvel An» 2000 tombera un vendredi 13, mais précise même que «le 1<sup>er</sup> janvier 2000 tombera un vendredi 13», l'explicitation complète de la date n'affectant pas l'effet obtenu.<sup>10</sup>

Le *traitement superficiel* de l'information (*shallow processing*) se produit lorsque le contenu d'une expression conceptuelle se trouve réduit à une information minimale qui permet de maintenir la présomption de pertinence (pour la notion de *shallow processing*, voir Carpenter & Just 1983; pour une discussion de son application au discours persuasif, voir Allott 2005). La fameuse «blague des survivants» est un exemple classique de traitement superficiel de l'information: on demande à la victime où doivent être enterrées les survivants d'un crash d'avion qui se produit exactement à la frontière entre deux pays (voir aussi Oswald 2010). Dans ce cas, l'analyse en «traitement superficiel» suggère qu'à l'expression *survivants* est associé, sur l'instant, et pour préserver la présomption de pertinence et donc la consistance qui lui est nécessaire, le concept de la catégorie superordonnée, *victime d'un crash* ou

8 Il avait déjà porté à notre connaissance cet exemple en 2005 lors d'une conversation informelle.

9 <http://www.philippevandel.com/2009/01/le-jour-de-l-an.html>

10 On ne s'agit évidemment pas d'une expérience scientifique, de sorte que le nombre de personnes ayant réagi à l'inconsistance n'est pas connu. Néanmoins, le fait que les nombreuses personnes montrées dans cette vidéo ne réagissent pas à l'inconsistance suffit à s'intéresser au phénomène. Par ailleurs, l'expérience de l'auteur et que même dans des contextes conversationnels, et même dans des contextes métadiscursifs, l'effet se produit en général aussi.

même *être humain*. Une modification conceptuelle est ainsi consentie inconsciemment et automatiquement sous la pression des attentes de consistance et de pertinence. L'*illusion cognitive* célèbre dite «illusion de Moïse» (*Moses illusion*), Reder & Kusbit (1991), est similaire: on demande combien d'animaux de chaque espèce Moïse embarque dans l'Arche. Dans d'autres cas encore, ce n'est pas la catégorie superordonnée, mais de simples connotations positives qui sont conservées – c'est du moins ce que suggère Allott (2005) qui analyse le sens de *démocratie* dans les énoncés d'officiels américains expliquant que les Etats-Unis entrent en guerre en Irak pour «promouvoir la démocratie»; selon Allott, la simple évocation des connotations positives suffit à amener au consentement.

Il convient toutefois de remarquer que les énoncés concernés portent des présuppositions, et que ce sont ces présuppositions qui sont affectées – sauf lorsque le traitement superficiel ne retient que des connotations. La blague des survivants présuppose l'existence de survivants; l'illusion de Moïse présuppose que Moïse a chargé l'Arche.

Il est possible que, comme le suggère Oswald (2010: 377-378), un tel processus de traitement superficiel se produise dans l'affaire du vendredi 13. Cela expliquerait élégamment pourquoi des éléments de l'arrière-plan sont laissés de côté, comme «des survivants sont vivants» ou «le Nouvel An est un 1<sup>er</sup> (janvier)». Dans leurs études fondatrices sur le rôle du contexte dans la manipulation discursive, Maillat & Oswald (2009, 2011) suggèrent que divers dispositifs discursifs conduisent le destinataire à opérer une contextualisation insuffisante des énoncés (notre terminologie) en l'amenant à laisser de côté des informations cruciales, qui donc resteront hors du contexte d'interprétation (lequel consiste en un ensemble de propositions permettant des inférences, selon Sperber & Wilson dont Maillat & Oswald s'inspirent), ouvrant la porte à des interprétations biaisées. Le traitement superficiel réalise un tel processus de tenir à l'écart des propositions qui seraient utiles pour une compréhension complète ou qui poseraient des problèmes de consistance avec le contenu propositionnel des énoncés concernés.

Il y a pourtant quelques éléments qui ne concordent pas avec une explication du phénomène «vendredi 13» qui soit uniquement fondée sur le traitement superficiel. En particulier, il y a ce fait étonnant: l'information *Le Nouvel An 2000 tombera un vendredi 13* se trouve suffisamment incorporée dans l'environnement cognitif du destinataire – fût-ce provisoirement – pour

qu'elle autorise des déductions supplémentaires *fondées sur sa vérité*. Cela constitue une différence significative avec l'histoire des survivants: la propriété d'être vivant ne peut plus être exploitée, on ne peut pas continuer l'histoire avec des individus qui ressuscitent. Avec le vendredi 13, au contraire, on observe une variété d'inférences possibles fondées sur le fait – parfaitement inconsistant – que le Nouvel An 2000 est un vendredi 13. Ainsi, aux personnes qui répondent qu'elles ne sont pas superstitieuses, Vandel demande confirmation qu'ils célébreront donc le réveillon tout à fait normalement, donc le *jeudi* 12. Cette dernière date n'est aucunement associée avec un jour superstitieux, et donc ne peut faire l'objet de l'accommodation prêtée à *vendredi* 13. Qu'on célèbre le réveillon le jeudi 12 est une inférence reposant sur la prémisse donnant vendredi 13 comme Premier de l'an, et non «un jour superstitieux» comme Premier de l'an. De la sorte, il ne s'agit pas seulement d'un traitement superficiel qui tient à l'écart l'inconsistance de l'information, mais aussi qu'une fois éliminée par d'autres informations données par l'énoncé (en l'occurrence que le Premier de l'an peut être un 13), ces dernières sont suffisamment solides pour autoriser d'autres inférences, tout aussi inconsistantes avec les informations d'arrière-plan.

Dans l'exemple du vendredi 13, de plus, on observe que le traitement de l'information, superficiel peut-être, n'amène pas à modifier le contenu pour un autre, plus large. Le traitement implique en même temps, d'une part, de ne pas traiter l'énoncé littéral dans ses implications sémantiques, tout en gardant, néanmoins – et c'est ce qui est surprenant –, ses implications comme utilisables dans des schémas inférentiels (comme c'est le cas pour les présuppositions en général). Dans le cas qui nous occupe, le journaliste commence par l'énonciation d'un fait (il a été calculé que...) puis pose une question qui présuppose le fait préalablement énoncé («Cela vous inquiète-t-il [que le Nouvel An 2000 tombe un vendredi 13]?»);<sup>11</sup> dès lors, en traitant la question, le fait est déjà traité comme une présupposition; cela n'aurait pas été le cas avec une question qui ne l'aurait pas présupposé, bien évidemment, par exemple «Cela vous semble-t-il possible?». Lors du traitement en direct de la question, l'évaluation critique est complexe à cause de l'exigence de rapidité de l'interaction vis-à-vis du temps demandé par une évaluation

11 On notera d'ailleurs que comme «il a été calculé X» implicite que X est vrai, le second énoncé amène donc une présupposition pragmatique à la Gazdar (1979) (qui n'en reste pas moins une présupposition *stricto sensu*).

critique; notre hypothèse est que la stratégie d'incorporer à des fins économiques, liées à la rapidité de l'énonciation et des réactions qu'elle suppose, des présuppositions posées par un locuteur, est d'abord une stratégie très utile en situation ordinaire (non machiavélique) d'interaction. L'information posée comme fait ne laisse pas le temps d'une évaluation mais se trouve reprise en présupposition par la question, cet emballage informationnel suffisant à activer le mécanisme d'incorporation à l'environnement cognitif de l'information concernée (donc: une accommodation de présupposition). Il faudrait pour l'empêcher des indices forts de son caractère douteux; or l'étonnant est que l'inconsistance logique ne suffise pas à être prise pour un tel indice, mais tel est le cas. Des paramètres externes viennent sans nul doute conforter l'interprétation: le journaliste est là avec tout son appareillage qui sert de garant de son sérieux professionnel et donc de la présomption de la pertinence de ses énoncés. Toutefois, ces paramètres ne sont pas explicatifs, du moins pas à eux seuls, tant bien d'autres présuppositions inconsistantes, notamment avec des faits perceptifs, ne sont pas automatiquement intégrées sur la seule foi de celui qui parle. De plus, l'effet «vendredi 13» est actif également dans des situations de communication parfaitement ordinaires, comme l'auteur a pu l'expérimenter. Dans le cas du vendredi 13, ce qui entre en jeu est que l'information est présentée comme une évidence qui est en dehors de la focalisation de la question.

Le cas du vendredi 13 n'a pas d'impact sur la vie sociale: il s'agit d'une «expérience» intuitive qui montre combien une information présupposée peut être robuste dans ses effets d'accommodation inconsistante. Notre second cas est différent à cet égard, tout en activant des procédés similaires, au niveau de granularité où nous nous trouvons: des propositions motivant l'énonciation, et donc indispensables à la pertinence, sont accommodées comme des présuppositions ordinaires. Il s'agit sans doute d'ailleurs d'un cas plus clair de présuppositions proprement discursives.

Le 29 novembre 2009, les électeurs suisses étaient appelés à voter pour ou contre l'interdiction de construire des minarets sur le territoire national; cet objet résultait d'une «initiative populaire», dispositif constitutionnel permettant de soumettre au vote une proposition ayant recueilli un nombre légal de signatures. En l'espèce, la proposition émanait d'un parti politique. Les citoyens ayant participé à la votation (car il y a traditionnellement un taux d'abstention élevé, qui se situe souvent autour des 40%) ont accepté la



proposition par plus de 57% de *oui*. De nombreuses pistes d'explication ont été proposées et relayées dans les médias. L'essentiel de ces explications, surtout fondées sur des considérations socioculturelles et psychosociales, suggéraient qu'il s'agissait d'un vote irrationnel, ancré avant tout dans la peur de l'étranger, le réflexe d'exclusion, l'impulsion de la protection ou du repli ethnique ou culturel.<sup>12</sup> Le manque de communication des autorités fédérales a également été pointé, à fort juste titre, bien que, pour des raisons sur lesquelles nous revenons plus bas, ce manque de communication puisse s'expliquer de manière assez pragmatique. Certes, sans un tel arrière-plan psychosocial ou socioculturel, le vote sur les minarets serait très étrange. Néanmoins, les médias n'ont pas accordé d'attention à un fait pourtant frappant, qui relève précisément de la confrontation entre la question soumise au vote, et les connaissances accessibles à l'arrière-plan, c'est-à-dire présentes dans l'environnement cognitif, des destinataires. Or, la question présupposait, à titre discursif, des informations fortement divergentes avec les connaissances de la plupart des destinataires; ces divergences eussent-elles été identifiées que le résultat aurait peut-être été bien différent. Et pourtant, comme avec le cas du *vendredi 13*, elles ne l'ont pas été.

Pour que l'objet soumis au vote soit pertinent,<sup>13</sup> un certain nombre de propositions devaient être vraies ou significativement probables:

- 12 Par exemple, Green (2010): «Les minarets incarnent la présence religieuse. Ils sont des incursions hautement visibles dans l'espace civil. Leur présence dans des paysages ruraux et urbains d'Europe brouille les frontières entre l'espace religieux et l'espace civil que les Européens ont travaillé à construire dans l'histoire moderne» (Green 2010: 621). Cette idée est certainement présente dans l'arrière-plan de nombreuses personnes, néanmoins la plupart du temps sous le seuil de la conscience. L'idée que les minarets sont les signes d'un Islam qui cherche à jouer un rôle dans la vie publique a même été soutenue par les promoteurs de l'interdiction (Green 2010). D'autres chercheurs ont suggéré l'hypothèse que l'interdiction des minarets avait quelque chose à voir avec une norme particulière de l'acceptabilité de l'immigration selon l'assimilation sociale (Gianni & Clavien 2012). Cependant ce qui nous concerne ici est de savoir comment la question a été comprise, et quel arrière-plan servant de contexte a dû être incorporé dans l'environnement cognitive des personnes afin d'en trouver un sens pertinent.
- 13 Une question référendaire en Suisse, y compris lors d'un référendum d'initiative populaire, prend traditionnellement la forme suivante: «Acceptez-vous l'initiative P», ici «Acceptez-vous l'initiative populaire «La construction des minarets est interdite?»» Le matériel fourni aux votants comprend divers fascicules exposant les positions des partis

- (a) Il y a un nombre pertinent de minarets existants ou projetés.
- (b) Les minarets pourraient modifier l'allure des paysages suisses (urbains ou ruraux).
- (c) Les minarets constituent une menace.

Dans le cadre des propositions que nous avons énoncées plus haut, (a), (b) et (c) constituent des présuppositions discursives, qui sont, en totalité ou non, admises pour que l'interdiction des minarets puisse avoir un sens pragmatique (quelle que soit la formulation choisie pour le scrutin, un fait sur lequel nous attirons d'ailleurs l'attention, car les présuppositions ci-dessus sont bien non linguistiques mais discursives).

La présupposition discursive (a) constitue une première motivation pour l'objet soumis au vote, car il serait peu pertinent de voter l'interdiction d'un type de construction par ailleurs totalement inexistant dans le pays (ce n'est cependant pas une présupposition existentielle au sens classique, car la question de l'interdiction de quelque chose ne présuppose pas *stricto sensu* son existence présente). La présupposition discursive a) est assez facile à accommoder, puisqu'il y a effectivement une communauté musulmane en Suisse et qu'il est rationnel de supposer qu'elle pratique sa religion dans des mosquées, et que les mosquées ont en général un minaret. Néanmoins, pratiquement aucun votant, hormis quelques privilégiés, n'ont vu un minaret de leur vie sur le territoire national: au moment du vote, il y en avait trois, tous situés dans des zones suburbaines. Néanmoins, ne jamais avoir rien vu ne forme pas une hypothèse d'inexistence, de sorte qu'il suffit de ne rien savoir pour que l'accommodation de (a) soit favorisée, de sorte qu'il n'y a pas ici d'effet réel d'inconsistance. La proposition (b) omet le fait – cette fois très largement connu – que les lois régissant l'aménagement du territoire sont drastiques, de sorte qu'il apparaît immédiatement fantaisiste, dès qu'on s'y attarde un instant, de supposer que des minarets pourraient apparaître à côté de Saint-Pierre à Genève, de la coupole du Palais Fédéral à Berne ou au milieu de chalets alpins. Il y a ici une vraie divergence, mais comme les lois et

et du gouvernement et leurs arguments. Ce qui nous importe ici n'est pas la forme linguistique, mais le fait qu'une telle question sur l'interdiction des minarets, en soi et indépendamment d'une forme linguistique particulière, mobilise des présuppositions discursives.

règlements sont sujets à évolution, le futur ne garantit pas la pérennisation de l'inconsistance présente, qui dès lors n'est plus un obstacle à la décision rationnelle, même s'il reste très improbable de voir jamais des minarets dans les villages des vallées helvétiques.<sup>14</sup> C'est bien la présupposition (c) qui retient plus sérieusement notre attention, car elle paraît de prime abord absurde, et pourtant nécessaire à tout vote d'interdiction des minarets: demander à interdire quelque chose présuppose, d'une manière purement pragmatique, la dangerosité, qui n'est qu'une forme particulière du non-souhaitable, et en même temps on ne voit pas comment un ornement architectural constituerait un danger. Bien entendu, c'est une métonymie qui va jouer le rôle de combler l'espace conceptuel entre le minaret et le danger, en pointant les utilisateurs potentiels de minarets – les musulmans – comme dangereux. Nul doute que ce transfert métonymique est favorisé, sinon qu'il a pour condition nécessaire, l'existence d'un lien dans l'imaginaire des individus, entre la religion musulmane et le danger, qui à son tour repose sur l'assimilation de l'islam à l'islamisme radical. Cet article n'a pas pour tâche de discuter ce point précis, mais il importe tout de même de signaler à quel point l'islam en Suisse est dépourvu d'attaches avec l'islamisme radical, et qu'on peut prendre pour signe de cette modération évidente le fait que les musulmans de Suisse n'ont jamais réclamé le moindre droit de construire des minarets. Quoi qu'il en soit, c'est donc par métonymie que le concept d'un danger des minarets s'est vu assimilé avec le danger de l'islamisme radical.

Il y a bien évidemment beaucoup à ajouter sur l'affaire des minarets, et l'analyse que nous venons de formuler n'a pas d'autre ambition que de rester partielle, concentrée sur des faits de traitement de l'information. L'important est que, dans les deux cas que nous avons abordés ci-dessus, de nouvelles propositions sont amenées sous la forme d'information d'arrière-plan présupposées, de telle manière que des informations effectivement connues et très assurées sont néanmoins occultées, sans même que cette occultation,

14 Une journaliste d'un grand quotidien suisse a rapporté à l'auteur l'anecdote suivante: une personne d'un niveau socioculturel élevé lui confiait, peu avant les votations, ne pas souhaiter voir surgir des minarets dans les paysages alpestres que traversait leur promenade. Cela montre que l'idée de *ne pas interdire* reviendrait à *tout autoriser*, ce qui relève d'un biais de polarité, auquel nous ne nous intéressons pas ici, mais qui est très pertinent pour comprendre l'événement. Voir Rigotti (2005) pour un développement sur le biais, ou «tentation», de polarité.

qui intervient pour maintenir la pertinence de l'énoncé, ne soit identifiée par le destinataire. Dans de tels cas, l'énoncé prend sa pertinence non pas grâce aux implicatures qu'il permet de dériver, mais essentiellement à cause de l'arrière-plan ou supposé tel dont l'incorporation est indispensable à la production de la pertinence. Cette pertinence est obtenue par cette incorporation dans l'arrière-plan. De la sorte, ces énoncés ont une *pertinence d'arrière-plan*, puisque ce sont les présuppositions discursives qui fournissent cette pertinence.

Il est logique de conclure qu'il s'agit d'assertions, au moins par leur fonction, par le biais de présuppositions, donc d'*assertions présuppositionnelles* ou d'*assertions d'arrière-plan*.

## 6. Conclusions et perspectives: une hiérarchie de biais cognitifs

Ne pas dépenser de l'effort et du temps au sujet de ce qui est déjà connu, partagé ou réputé tel, c'est-à-dire à propos de savoir déjà stabilisé, est une propriété de la cognition qui n'est pas inattendue et qui converge avec l'orientation vers la maximisation de pertinence (Sperber & Wilson 1995) ou toute autre orientation impliquant un seuil d'équilibre entre contraintes. Rien d'étonnant donc, dans cette perspective, que les propriétés pragmatiques du traitement des présuppositions que nous avons examinées en font des instruments heuristiques très efficaces dans les situations ordinaires de communication. Mais ceci ne va pas, comme on l'a vu, sans former un espace problématique d'absence de traitement critique, évaluatif, de l'information qu'elles contiennent, permettant à un locuteur une exploitation intentionnelle (consciente ou non), afin d'amener une information dans le réseau de croyances de l'interlocuteur en contournant le filtre de la présomption de pertinence.

Nous avons suggéré plus haut que la puissance des présuppositions en général et des présuppositions discursives en particulier a à voir avec le statut qu'elles exhibent d'appartenir à de l'information ancienne, même si tel n'est pas le cas. Cela permet de rapprocher les phénomènes que nous avons

observés d'un grand nombre de biais cognitifs largement documentés dans la littérature, mais qui ont été décrits sans qu'il soit observé qu'ils ont un lien direct avec la nature présuppositionnelle des informations qu'ils exploitent. Si cette analyse est exacte, alors il faut conclure que le biais présuppositionnel que nous avons observé ci-dessus se manifeste comme fondement d'un ensemble de biais déjà connus.

Les biais de cadrage (*framing biases*) (Tversky & Kahneman 1974) jouent sur un cadrage qui peut être compris comme un ensemble d'informations d'arrière-plan, donc qui joue le rôle de présuppositions, qui s'activent lors du traitement d'assertions ou de questions subséquentes. Un exemple classique de Tversky & Kahneman (1974) est la question suivante: *Le pourcentage de nations africaines à l'ONU est-il supérieur ou inférieur à 10%?* (le pourcentage réel dépasse largement les 50%). Les sujets sont invités à répondre par une estimation, et il apparaît que ces dernières sont proches des 10%. Il est raisonnable de considérer que la question introduit une présupposition discursive, à savoir que le pourcentage est *autour de 10%*, ce qui donne de la pertinence au nombre 10, sous l'hypothèse que s'il est mentionné, c'est qu'il méritait d'être mentionné, donc qu'il est pertinent. Des effets connus d'étiquetage (*labelling effects*), et de mésinformation (*misinformation effects*), étudiés par Loftus & Palmer (1974) et Loftus & Burns (1978), associent à des mots de l'information encyclopédique présupposée ou activent des présuppositions existentielles, typiquement avec des descriptions définies. Un effet d'étiquetage survient par exemple lorsqu'on donne un nom erroné à un objet, par exemple en parlant d'un panneau «stop» sur une photo pour désigner un panneau «cédez la priorité», et que cela amène un individu à croire sur le long terme que le panneau était bien un stop si le fait est amené par une question qui le présuppose. Un effet de mésinformation survient par exemple lorsqu'une évaluation, par exemple de la vitesse de deux voitures ayant eu un accident, est affectée par le verbe utilisé dans la question, par exemple «à quelle vitesse les voitures roulaient lorsqu'elles se sont heurtées (*«hit»*) / lorsqu'elles se sont percutées (*«smashed»*)». Dans tous ces cas, il s'agit d'effets présuppositionnels.

On ne peut éviter de comparer en outre les biais présuppositionnels et le biais de confirmation; cela conduit à une hypothèse supplémentaire et plus forte, qui doit encore être explorée. Si le fait de *confirmer une hypothèse déjà entretenue* consomme moins de carbone que de *contredire une hypothèse déjà*

*entretenu*, alors l'information préexistante dispose d'une force d'ancrage directement liée aux préférences énergétiques du cerveau. Il est plausible, bien que cela doive faire l'objet d'études, que la force d'ancrage des présuppositions soit directement liée à cette préférence énergétique. Cela fait d'autant plus sens si on considère les présuppositions sous les angles combinés de i) leur pertinence, ii) l'absence d'évaluation critique, qui consomme de l'énergie, et iii) leur force d'ancrage dans l'environnement cognitif. Il est envisageable que, dans cette perspective, le biais présuppositionnel, qui se manifeste en particulier dans les présuppositions discursives, mais fonderait également les biais d'étiquetage, de cadrage, de mésinformation – entre autres – soit à son tour une simple manifestation d'une propriété de beaucoup plus bas niveau, qui a à voir avec le statut économique d'une information ancienne et le coût de son infirmation, donc le biais de confirmation.

Dans cet article, nous avons proposé d'étendre au discours la notion de présupposition, généralement conçue comme participant au sens de l'énoncé, soit au niveau sémantique soit au niveau pragmatique. Notre proposition consiste à considérer comme des présuppositions particulières les informations qui doivent être mobilisées pour qu'un énoncé non seulement puisse faire sens, mais puisse avoir un sens *pertinent* dans les circonstances. Nous avons suggéré que ces informations, les présuppositions discursives, ont une force persuasive particulière, car elles se présentent comme anciennes (même si elles ne le sont pas en l'espèce) et peuvent échapper à la vigilance épistémique (selon le terme de Sperber & *al.* 2010). Néanmoins, nous soulignons ceci: ces faits de traitement de l'information se réalisent par des mécanismes mentaux automatiques, qui forment la partie stable du processus de persuasion par présupposition que nous avons abordé; ces processus réguliers amènent un gain substantiel dans la communication coopérative (dans un contexte non machiavélien), à savoir qu'il est inutile de traiter une information préalable, puisque si elle est préalable, elle a déjà été traitée.

Sperber & Wilson (1995: 217) séparent *type de contenu* et *plan* (avant- ou arrière-plan), précisant que c'est au cours du traitement de l'information, guidé par la recherche de pertinence, que l'interlocuteur la considérera comme *avant-* ou *arrière-plan*. Avec des accommodations présuppositionnelles du type de celles que nous venons d'observer, qu'il s'agisse de présuppositions *stricto sensu* ou de ces riches présuppositions discursives, nous

obtenons un croisement inattendu entre un haut degré de pertinence et un classement dans l'arrière-plan, avec pour résultat que cette pertinence modifie significativement l'environnement cognitif du destinataire sans que des effets d'étrangeté ou d'inconsistance soient (facilement) perçus. Il s'agit donc de *pertinence d'arrière-plan*.<sup>15</sup>

## Bibliographie

- Abrusan, M. (2013): «A note on quasi-presuppositions and focus», *Journal of semantics* 30, 257-265.
- Allott, N. (2005): «The role of misused concepts in manufacturing consent: A cognitive account», in Saussure, L. de & Schulz, P. (éds), *Manipulation and Ideologies in the XXth century: Discourse, Language, Mind*. John Benjamins, Amsterdam/Philadelphia, 147-168.
- Amaral, P., Cummins, C. & Katsos, N. (2011): «Experimental evidence on the distinction between foregrounded and backgrounded meaning», *Proceedings of the ESSLLI 2011 workshop on projective content*, ms, 1-7.
- Carpenter, P. A. & Just, M. A. (1983): «What your eyes do when your mind is reading», in Rayner, K. (éd.), *Eye movements in reading: Perceptual and language processes*. Academic Press, New York, 275-307.
- Carston, R. (1998): «Negation, 'presupposition' and the semantics/pragmatics distinction», *Journal of linguistics* 34 (2), 309-350.
- Clark, H., Schreuder, R. & Buttrick, S. (1983): «Common ground and the understanding of demonstrative reference», *Journal of Verbal Learning and Verbal Behavior* 22, 245-258.
- Ducrot, O. (1972): *Dire et ne pas dire*. Hermann, Paris.
- (1980): *Le dire et le dit*. Minuit, Paris.
- Dijk, T. A. van (1998): *Ideology. A multidisciplinary approach*. Sage, London.
- (2011): «Discourse, knowledge, power and politics: Towards critical epistemic discourse analysis», in Hart, C. (éd.), *Critical Discourse Studies in Context and Cognition*. John Benjamins, Amsterdam/Philadelphia, 27-64.
- Fintel, K. von (2000): *What is presupposition accommodation?* ms., MIT, Cambridge. <http://web.mit.edu/fintel/www/accomm.pdf>.
- (2008): «What is presupposition accommodation, again?», *Philosophical perspectives* 22 (1), 137-170.

15 *Background relevance* dans Saussure (2013). Une telle pertinence d'arrière-plan converge avec les observations de Sandford & Sturt (2002) et Amaral & al. (2011) sur le caractère plus superficiel du traitement des informations d'arrière-plan.

- Gazdar, G. (1979): *Pragmatics: Implicature, Presupposition, and Logical Form*. Academic Press, New York.
- Gianni, M. & Clavier, B., (2012): «Representing gender, representing Muslims? Gender and figures of otherness in public discourses in Switzerland», in Flood, C., Hutchings, S., Miazhevich, G. & Nickels, H. (éds), *Political and cultural representations of Muslims: Islam in the plural*. Brill, Leiden, 113-130.
- Green, T. (2010): «The resistance to minarets in Europe», *Journal of Church and State* 52 (4), 619-643.
- Greco, S. (2003): «When presupposing becomes dangerous. How the procedure of presuppositional accommodation can be exploited in manipulative discourses», *Studies in Communication Sciences* 3 (2), 217-234.
- Hart C. (2011): «Legitimizing assertions and the logico-rhetorical module: Evidence and epistemic vigilance in media discourse on immigration», *Discourse Studies* 13 (6), 751-769.
- Haviland, S. & Clark, H. (1974): «What's new? Acquiring New information as a process in comprehension», *Journal of Verbal Learning and Verbal Behavior*, 13 (5), 512-521.
- Heim, I. (1983): «On the projection problem for presuppositions», *West Coast Conference on Formal Linguistics* 2, 114-125.
- Karttunen, L. (1974): «Presupposition and linguistic context», *Theoretical Linguistics* 1, 181-193.
- Kleiber, G. (2012): «Sur la presupposition», *Langages* 186, 21-36.
- Lewis, D. (1979): «Scorekeeping in a language game», *Journal of philosophical logic* 8, 339-495.
- Loftus, E. F., Miller, D. G. & Burns, H. J. (1978): «Semantic integration of verbal information into a visual memory», *Human Learning and Memory* 4, 19-31.
- Loftus, E. F., Palmer, J. C. (1974): «Reconstruction of Automobile Destruction: An Example of the Interaction Between Language and Memory», *Journal of Verbal Learning and Verbal Behavior* 13, 585-589.
- Macagno, F. (2012). «Reconstructing and assessing the conditions of meaningfulness. An argumentative approach to presupposition», in Ribeiro, H. (éd.), *Inside Arguments: Logic and the Study of Argumentation*. Cambridge Scholars Publishing, Newcastle upon Tyne, 247-268.
- Maillat, D. & Oswald, S. (2009): «Defining manipulative discourse: the pragmatics of cognitive illusions», *International Review of Pragmatics* 1(2), 348-370.
- (2011): «Constraining context: a pragmatic account of cognitive manipulation», in Hart, C. (éd.), *Critical discourse studies in context and cognition*. John Benjamins, Amsterdam/Philadelphia, 65-80.
- Oswald, S. (2010): *Pragmatics of uncooperative and manipulative communication*. PhD thesis. University of Neuchâtel, ms.
- Petty, R. E. & Cacioppo, J. T. (1986): *Communication and Persuasion: Central and Peripheral Routes to Attitude Change*. Springer-Verlag, New York.
- Pohl, R. (2004): *Cognitive Illusions: A Handbook on Fallacies and Biases in Thinking, Judgement and Memory*. Psychology Press, London.
- Reder, L. M. & Kusbit, G. W. (1991): «Locus of the Moses illusion: Imperfect encoding, retrieval, or match?», *Journal of memory and language* 30, 385-406.
- Rigotti, E. (2005): «Towards a typology of manipulative processes», in Saussure, L. de & Schulz, P. (éds), *Manipulation and Ideologies in the XXth century: Discourse, Language, Mind*. John Benjamins, Amsterdam/Philadelphia, 61-84.



- Rigotti, E. & Rocci, A. (2001): «Sens, non-sens, contresens. Tentative d'une définition explicative», *Studies in communication science* 1, 45-80.
- Rocci, A. (2005): «Are manipulative texts 'coherent'? Manipulation, presuppositions and (in-)congruity», in Saussure, L. de & Schulz, P. (éds), *Manipulation and Ideologies in the XXth century: Discourse, Language, Mind*. John Benjamins, Amsterdam/Philadelphia, 85-112.
- Sandt, R. van der (1992): «Presupposition projection as anaphora resolution», *Journal of semantics* 9, 332-377.
- Saussure, L. de (2013): «Background Relevance», *Journal of Pragmatics* 59 (Part B), 178-189.
- Sanford, A. & Sturt, P. (2002): «Depth of processing in language comprehension. Not noticing the evidence», *Trends in cognitive sciences* 6 (9), 382-386.
- Seuren, P. (2000): «Presuppositions, negation and trivalence», *Journal of linguistics* 36, 261-297.
- Simons, M. (2005): «Presuppositions and relevance», in Szabo, Z. (éd.), *Semantics versus Pragmatics*. Oxford University Press, Oxford, 329-355.
- (2006): *Presupposition without common ground*. Carnegie Mellon University, ms.  
[http://www.semanticsarchive.net/Archive/TY3NTBmY/Presupposition%20without%20Common%20Ground\\_Simons.pdf](http://www.semanticsarchive.net/Archive/TY3NTBmY/Presupposition%20without%20Common%20Ground_Simons.pdf)
- Soames, S. (1982): «How presuppositions are inherited: A solution to the projection problem», *Linguistic Inquiry* 13, 483-545.
- Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G. & Wilson, D. (2010): «Epistemic vigilance», *Mind and Language* 25 (4), 359-393.
- Sperber, D. & Wilson, D. (1995) [1986]. *Relevance: Communication and cognition*. Blackwell, Oxford.
- Stalnaker, R. (1974): «Pragmatic Presuppositions», in Munitz, M. & Unger, P. (éds), *Semantics and Philosophy*. New York University Press, New York, 197-213.
- Stalnaker, R. (1998): «On the representation of context», *Journal of Logic, Language and Information* 7 (1), 3-19.
- (2002): «Common ground», *Linguistics and Philosophy* 25, 701-721.
- Thomason, R.H. (1990): «Accommodation, meaning, and implicature: Interdisciplinary foundations for pragmatics», in Cohen, P., Morgan, J. & Pollack, M. (éds), *Intentions in communication*. MIT Press, Cambridge, 325-363.
- Thomason, R.H., Stone, M. & DeVault, D. (2006): «Enlightened update: A computational architecture for presupposition and other pragmatic phenomena», *Paper presented at the Workshop on Presupposition Accommodation*, Ohio State University.  
<http://www.pragmatics.osu.edu/links/events/enlightened-update.pdf>
- Tversky, A. & Kahneman, D. (1974): «Judgment under uncertainty: Heuristics and biases», *Science* 185 (4157), 1124-1131.
- Walton, D. (1999): «The fallacy of many questions: on the notions of complexity, loadedness and unfair entrapment in interrogative theory», *Argumentation* 13, 379-383.



# Pragmatics, cognitive heuristics and the straw man fallacy

Steve OSWALD, University of Fribourg & University of Neuchâtel  
Marcin LEWIŃSKI, Universidade Nova de Lisboa

## 1. Introduction

In the literature on argumentation, the straw man fallacy denotes the misrepresentation of someone's position in order to easily refute that position (see Aikin & Casey 2011, Lewiński 2011, Bizer, Kozak, & Holterman 2009, van Eemeren & Grootendorst 1992: Ch. 11, van Laar 2008, Talisse & Aikin 2006, Walton 1996, Walton & Macagno 2010). An argumentative move that involves a straw man can thus be characterised by its two constitutive elements (Lewiński 2011): the function of refuting an opponent's position and the form consisting in various methods of misrepresentation of the original position (misquotation, selective quotation, taking out of context, attacking a fictitious opponent, etc.). The straw man fallacy has drawn considerable scholarly attention, not least because of its prevalence in public discourse (see esp. Aikin & Casey 2011 and Talisse & Aikin 2006).

What is a remarkable but often overlooked feature of the straw man is that, contrary to most other common fallacies such as *ad hominem* or *ad baculum*, it is a *meta*-discursive fallacy. It “operates” on someone else's discourse that serves as material for linguistic manoeuvring.<sup>1</sup> It therefore does not neatly fit into the Aristotelian categories of verbal (“dependent on language”) and material (“not dependent on language”) fallacies (see

1 As mentioned above, this includes purported discourse attributed to some fictitious or unidentified/unidentifiable “author”: frequent examples include clauses such as “People out there may say P”, “It is well-known that P”, etc. In this variant, the question of who exactly these “people” are becomes crucial.

Hamblin 1970).<sup>2</sup> A straw man attack involves a certain unreasonable, sometimes manipulative, re-interpretation of another position. Hence, it plays on the interpretative resources of arguers and its identification as such requires careful meta-linguistic (or meta-discursive) attention.

In our previous work (Lewiński & Oswald 2013) we argued that to fully grasp both the pragmatic and the argumentative mechanisms behind the straw man fallacy, one has to answer two questions: (i) *when* can we justifiably say that the straw man occurred? and (ii) *how* can we explain its remaining covert and thus its persuasive or even manipulative potential? We posited that a satisfactory answer requires a theoretical combination of two branches of pragmatic inquiry: a *normative* pragmatic theory of argumentation, such as Pragma-Dialectics (henceforth PD), and a *cognitive* pragmatic model of meaning (and in particular of the cognitive mechanisms governing interpretation), such as the one offered by Relevance Theory (henceforth RT). Our rationale for such combination was that the normative pragmatic component of PD provides a framework for answering question (i), while RT's empirically adequate cognitive pragmatic account is capable of addressing the problems involved in question (ii). We found these two theories complementary in their goals and methods for the *analysis* of argumentative discourse. We argued that this is clear in both their pragmatic take on meaning in discourse and their heavy reliance on the notion of context. Within this hybrid framework, we stipulated contextual, pragma-dialectical conditions for the identification of the alleged straw man attacks and provided a basic cognitive account of the manipulative effectiveness of straw man abuses (see also section 3.2).

This contribution builds on our previous results. Its goal is twofold: first, we aim to refine the grounds on which we claim both theories can be combined in argumentative, and more specifically rhetorical, analysis and, second, extend our cognitive account, so that question (ii) above becomes more prominent than question (i) which was primarily addressed in our previous research. We will thus carefully explore the pragmatic relation between the normative violation that makes a criticism fallacious and the

2 It may be taken as a conversational sub-category of *ignoratio elenchi* of sorts, and thus as a fallacy not dependent on language. However, it clearly depends on linguistic means of misrepresentation, which would suggest an opposite classification. Note that neither the straw man nor “ad” fallacies were discussed by Aristotle (see Hamblin 1970).

cognitive “mistake” that makes an unreasonable straw man acceptable to reasonable hearers. More specifically, we will investigate the cognitive mechanisms responsible for an addressee’s failure to identify the straw man as an unreasonable, fallacious, critique.

To address these questions, we will proceed as follows. First, in section 2 we will critically review the most relevant attempts within Argumentation Theory to account for the treacherousness of fallacies in terms of simplified cognitive heuristics and elaborate on the cognitive underpinnings of argument processing following recent findings in the cognitive study of epistemic aspects of communication (see Sperber *et al.* 2010). Second, in section 3 we will highlight some issues raised by Walton’s account (2010) and outline what we take to be a more fruitful argumentative analysis of the straw man fallacy, by combining PD and RT into a consistent framework. The crucial element of our proposal is an elaborate cognitive account of the pragmatic mechanisms involved in the hearer’s interpretation of fallacious straw man critiques. Finally, section 4 will illustrate the working of our proposal through the analysis of an example taken from Swiss political discourse. We will try to detail the subtleties of a naïve, speaker-to-speaker interpretation, consisting of context-bound and implicit meanings and conclude on the relationship between understanding and accepting. In plain words, we will be concerned with the relationship between what a speaker means and what the addressee ends up accepting.

## 2. Argumentative fallacies and cognitive heuristics

### 2.1 *Heuristics and biases in Argumentation Theory*

The idea of gaining a fuller account of fallacies by “bring[ing] the normative dimension better into relation with the psychological dimension” (Walton 2010: 160) is not new, and indeed it would be surprising if it were. Fallacies, in the end, while being discursively defined as misuses of argumentation, can also be regarded from a cognitive perspective as bias-generated errors in inferential processes of reasoning and judgment that we make ourselves (cf.

Correia 2011) or unwittingly accept when others make them. Such cognitive “errors” in our view have to be regarded as information-processing paths yielding outputs that would have been different – in particular for the epistemic sake of the cognitive system – in case the individual had been led to mobilise critical information. The nature of the error does therefore not lie in an inherently flawed inferential procedure but rather in the reliance on a heuristic procedure solicited in a situation that would be more advantageously dealt with by carefully considering additional critical information. Under this view, producing or failing to spot a fallacy is not a matter of doing something inherently “wrong”, but rather the result of failing to select the most optimal strategy given the circumstances. Whether this situation was deliberately aimed at by the speaker or not has no incidence on the cognitive characteristics of the fallacious move in terms of its processing. A genuinely cognitive account of fallacies should consequently not only explain how these biases work, but also specify the conditions under which they operate and turn out to be argumentatively and epistemically disadvantageous.<sup>3</sup> We believe that part of the answer to these questions follows from the consequences of one of the defining features of fallacies, namely their covert nature. Fallacies are rhetorically effective, we contend, because they are not recognised as such, which is a property they share with deception (see Maillat & Oswald 2009, 2011, Oswald 2011, Oswald forth.).

Despite the manifest relevance of cognitive insights for the study of argumentation, research on cognitive aspects of reasoning (and by extension those of argumentation) has traditionally stayed within the bounds of cognitive psychology since Wason’s seminal works on logical reasoning in the 1960’s (Wason 1960, 1966) and Kahneman & Tversky’s pioneering work

3 A reviewer rightly pointed out that the notion of argumentative error should be carefully dealt with. It makes sense to speak of error only when a norm against which specific phenomena can be measured exists. The normative component of our account precisely allows us to judge whether arguments are fallacious or not, i.e., whether they amount to errors or not. On the cognitive side, the type of cognitive processing fallacies induce is not erroneous *per se*, to the extent that heuristics habitually perform well and yield good enough results, but suboptimal if we consider that the most cautious, reasonable and thus the most advantageous response in argumentation is the adoption of critical and reasonable doubt.

on cognitive heuristics (Kahneman & Tversky 1972, 1973).<sup>4</sup> Yet, some attempts to incorporate the findings of this body of research have been made in the study of argumentation (e.g., Jackson 1996, Walton 2010), and it is to the discussion of these that we now turn.<sup>5</sup>

Jackson (1996) argues that the treatment of fallacies delineated by the question of “what is wrong with certain patterns of argument” should be complemented by “serious attention to what is persuasive about these patterns – why, given that they are defective, they often gain assent” (Jackson 1996: 101). She develops an account of the persuasiveness of fallacies from the perspective of her concept of conversational argument, by making use of empirical results in persuasion studies and social psychology. According to her, much of ordinary conversation is governed by “a broad and unremarked presumption of acceptability” (1996: 111) directed towards our interlocutors’ discourse. Jackson argues that we simply tend to accept what others claim, unless there are some important overriding reasons that lead us to challenge a given claim, such as inconsistency of the claim with what we already believe, or suspicions about the speaker’s trustworthiness and competence (see below for an elaboration on this very point). Motivated by such concerns, we take to examine the claim and its justification. The depth of the examination depends on factors such as our interest in the issue and our prior knowledge of it. Importantly, we stop our examination of (potentially fallacious) argumentation “whenever the cost of careful thinking exceeds what [we] think [...] the issue is worth” (Jackson 1996: 104) – in other words, we stop critical examination of a claim and its justification once we no longer expect it to be relevant to us. For such reasons, we tend to economise our assessment of argumentation by deploying simplified heuristics, or shortcuts of reasoning, that are often fast and frugal but not deeply critical (see Tversky & Kahneman 1974). As a result, we may be misled into accepting as sound a piece of argumentation that would end up being regarded as fallacious under closer scrutiny. In this sense, fallacies are treacherous because they are

4 See also Evans (2004), Evans & Frankish (2009), Gigerenzer *et al.* (2011) for state-of-the-art contributions on the topic.

5 It should nevertheless be noted that the links between argumentation and cognition are becoming increasingly salient in contemporary scholarly publications (e.g. Mercier & Sperber 2009, 2011, Mercier 2012, Bardone 2011, Maillat 2013, Correia 2011, Oswald 2011).

“incidental products of evaluation heuristics that can be given good defence as diagnostic tools” (Jackson 1996: 101). Jackson illustrates the working of her concepts by analysing fallacious appeals to authority and invalid forms of categorical syllogisms.

We find Jackson’s account very useful in terms of its reliance on a cognitive psychological body of research, but not satisfactorily elaborated for two main reasons: i) her particular conception of fallacies seems to miss a crucial normative dimension and ii) her account can be further grounded in light of recent findings in cognitive approaches to psychology and anthropology.

First, from our perspective, Jackson seems to have gone too far in claiming that fallacies lie, basically, in the eye of the beholder: “informal fallacies may describe not the materials presented by a speaker but the interpretive and reconstructive choices of the hearer” (Jackson 1996: 105). While we recognise that the second half of this characterisation is useful to address a research question concerned with a rhetorical account of *how* it is that fallacies may actually be effective, we believe that the first part of her formulation neglects the idea that fallacies are above all transgressive argumentative moves *realised in discourse* – i.e., they do lie in the “materials presented by a speaker” (*ibid.*). Accordingly, the study of fallacies equally needs to develop along a normative dimension, one which identifies clear criteria to discriminate sound from fallacious arguments. The identification of these criteria, we contend, constitutes a necessary step for the explanation of why fallacies in general can be defined as moves which violate (dialectical, logical) norms of reasonable argumentation, and why a given argument is fallacious in particular; this seems to be missing from Jackson’s account. Second, Jackson draws on the results of persuasion studies which are based on behavioural data but which do not explain exactly how the cognitive heuristics responsible for the success of fallacious arguments may actually work. We will turn to discuss research that frontally addresses this last issue in section 2.2.

Similarly to Jackson, Walton (2010) explores the possibility of elucidating the deceptive character of many informal fallacies of argumentation in terms



of their connection to cognitive heuristics.<sup>6</sup> He does so by relying on the concepts developed in his theory of argumentation schemes. A heuristic, thus, is “a mediating concept between the notion of fallacy and defeasible argumentation scheme” (Walton 2010: 160). To explain this mediating function, Walton introduces the notion of a *parascheme*: “a device that can be used to represent the structure of a heuristic as a speedy form of inference that instinctively jumps to a conclusion and is commonly used to make decisions” (Walton 2010: 163). Paraschemes, which are schematic discursive devices used to represent the structure of argumentative heuristics, are argumentation schemes stripped of all the “prerequisites” that make a conclusion reached by a proper argumentation scheme a presumptively reasonable result of informal reasoning. These omitted prerequisites include additional premises, assumptions, rebutted exceptions, successfully answered critical questions, etc. Therefore, by relying on mere paraschemes, heuristics generate simplified and crude forms of argument prone to various errors and abuses that result in fallacies. Yet, these simplified forms are often astonishingly effective methods of reaching a conclusion, and thus they may “seem valid, but are not” (Walton 2010: 160, echoing the “standard” definition of fallacies – see Hamblin 1970: 12). According to Walton, here lies the treacherous character of many fallacies, especially those that are clearly based on the incorrect use of an argumentation scheme: *argumentum ad verecundiam* (argument from expert opinion), *argumentum ad ignorantiam* (argument from lack of knowledge), or *argumentum ad baculum* (argument from threat or bad consequences).

Walton’s account posits the existence of argumentative heuristics but does not draw on recent cognitive psychological research in support of his views. As a consequence, although the idea is very much appealing, it remains to be seen whether these dedicated paraschemes have cognitive plausibility at all, or if they rely on more general heuristics that may be used to attend to specific types of arguments when necessary. In other words, the specific role and mode of intervention of these heuristics is still unclear, from a cognitive perspective, in Walton’s account.

6 Walton seems to be unaware of Jackson’s research, or at least so is suggested by the absence of Jackson’s work in his discussion of the literature and in his reference list.

## 2.2 *Epistemic vigilance: cognitive insights into argument processing*

Jackson conceives of fallacies as “failed diagnostic strategies” (1996: 111). Although she presents this conception as an attempt to characterise what fallacies are, we believe, as pointed out above, that this provides only one half of a comprehensive account of fallacious argumentation, as the focus is here on how fallacies work from a cognitive perspective, and not on what fallacies are, *compared to sound arguments*. Still, we think that the direction of research she delineates – which is more in tune with rhetorical concerns such as argumentative effectiveness<sup>7</sup> – is worth pursuing by mobilising insights from contemporary cognitive science. This sub-section will accordingly deal with cognitive constraints that may influence the persuasive success of fallacious argumentation.

There are indeed nowadays serious and promising grounds to start examining, within the framework of argumentation theory, Jackson’s characterisation of fallacies as failed diagnostics and assessing how they work from a cognitive perspective; for this we need to consider the cognitive constraints that can (mis)lead argument processing by recipients so as to drive their attention away from the fallacious nature of the argument. What we are therefore concerned with here is how fallacies manage to prevent their recipients from engaging in any sort of critical thinking with respect to the issue at stake. What in other academic circles is referred to as critical thinking has very recently started to attract attention in cognitive psychology and anthropology through the works of Sperber and his colleagues, who define our critical abilities in terms of epistemic vigilance (see Sperber *et al.* 2010 and its translation in this volume).

Sperber *et al.* posit that humans “have a suite of cognitive mechanisms for epistemic vigilance, targeted at the risk of being misinformed by others” (2010: 359). These cognitive filters are taken to monitor incoming information at all times, calibrating trust in its source while simultaneously

7 The notion of argumentative effectiveness will be here understood as an argument’s ability to bring about the desired perlocutionary effect in the addressee (i.e., being convinced, persuaded, etc.). For this reason, whether an argument is sound or fallacious is irrelevant to its qualification in terms of effectiveness. We also note that in this chapter the terms “effective” and “successful” will be used interchangeably to qualify arguments that have fulfilled this perlocutionary purpose.

assessing message consistency. The two dimensions (speaker reliability / trustworthiness and message consistency – both also mentioned by Jackson) are extremely relevant in particular when we consider the case of argumentative discourse, which is verbally manifested in communication as sets of propositions exhibiting a link of justification. Both sound and fallacious arguments should therefore be expected to try to satisfy epistemic vigilance requirements in order to promote their conclusions' perceived acceptability. Along these lines, it seems reasonable to assume that recipients of argumentation take into account at least the following two dimensions before they reach the stage where they are convinced (or unconvinced): i) the amount of trust they credit the speaker with and ii) the quality – in terms of consistency – of the information that is brought to their attention. The twofold idea behind epistemic vigilance, to put it crudely, is obviously that we tend to be more convinced by people we trust and messages we find consistent and truthful than by dodgy people and ill-evidenced or dubious, perhaps contradictory, messages. Fallacies take every chance to secure an apparent display of the characteristics of sound and acceptable arguments (i.e., trustworthiness and consistency). An inquiry into this phenomenon accordingly requires that we pay closer attention at trust and message consistency in communication.

### 2.2.1 *Assessing trust (and exploiting the way we assess trust)*

How can we defend, from a cognitive perspective, the assumption that fallacies may strive to remain undetected by exploiting trust? This first type of constraint could be taken to directly follow from what some philosophers define as a natural tendency not to be systematically sceptical in communicative settings, or, in other words, a tendency to regard as true – and believable – what others tell us (which is also what Jackson has in mind). This claim is fairly widespread in philosophical circles and can find its roots in Thomas Reid's conception of epistemic trust as a default disposition (cf. Reid 2000 [1794]). It is echoed nowadays in Ruth Millikan's work, who considers communication to be some kind of trustworthy perception by proxy (cf. Millikan 1987) and in Burge's take on communication, who considers in his *Acceptance principle* that we "accept as true something that is presented as true and that is intelligible [to us] unless there are stronger reasons not to do so" (Burge 1993: 467). Within contemporary pragmatics,

this very idea also underlies Grice's conversational maxim of quality, which enjoins conversational participants to provide information they consider themselves true and to refrain from contributing information they lack evidence for (see Grice 1989: 27).

This is obviously not to say that human beings are systematically and hopelessly gullible creatures; rather, this is a characterisation of a tendency to believe *in the absence of further evidence to the contrary*. The story nevertheless seems to be more complex, because there are grounds to assume that absence of further evidence to the contrary is the result of some sort of cognitive filtering mechanism.

Sperber *et al.* contend in this respect that what seems at first hand to correspond to default gullibility is not tantamount to blind trust; trust is not there by default as a precondition of communication, but it is mediated by epistemic vigilance filters; according to them, “[w]e could not be mutually trustful *unless* we were mutually vigilant” (Sperber *et al.* 2010: 364). Epistemic vigilance so described is thus partly directed at assessing whether we can trust the person who is informing or trying to convince us of something. These mechanisms are unconscious, to a large extent intuitive, and are supposed to alert us whenever there are grounds to suspect malevolent or deceptive intentions on behalf of the speaker. In other words, they are responsible, among other things, for checking whether there is available evidence not to trust a speaker.

Now, from the perspective of fallacious argumentation, it is reasonable to assume that a “successful” fallacy will manage to satisfy expectations of epistemic vigilance, so that no “red flag” is raised in the cognitive system as it is processed. And as far as monitoring the trustworthiness of the source is concerned, the fallacious argument will consequently precisely need to display every assurance that the speaker *is* trustworthy. In information-processing terms, considering someone to be trustworthy has important consequences on the epistemic features of whatever this person utters: crucially, we will deem what trustworthy people tell us to be epistemically strong – i.e., to reliably enhance our knowledge of the world. Managing to “discursively” establish a speaker's trustworthiness can therefore prove to be a way of managing to convey epistemically strong information, as long as it is said to stem from such a speaker. In argumentation parlance, we are therefore tackling speaker's *ethos*, namely how people perceive her in terms

of competence, benevolence, expertise, reliability, trustworthiness, etc. If we look at the traditional list of fallacies, the so-called “gang of eighteen” (see Woods 2004), one of them in particular seems to be exactly tailored to achieve this effect: the argument *ad verecundiam*, or appeal to the authority of expert judgment (cf. Walton 1989, Pilgram 2012).<sup>8</sup> By launching an appeal to a (secretly) inadequate and irrelevant expert, the speaker is expecting the claim brought forth to inherit its epistemic strength from the figure she presents as an expert. If the appeal is successful, it means that the proposition associated with it enjoys a heightened epistemic status; what is more, if the addressee fails to see that the expert figure mentioned is inadequate, epistemic vigilance requirements might even be fulfilled, leaving the fallacy unthreatened as no reasons to question its validity have arisen.

### 2.2.2 *Assessing message consistency (and exploiting the way we assess message consistency)*

Alongside vigilance towards the source, epistemic filters are also vigilant towards the content of the communicated message. This means that we are naturally equipped to identify inconsistencies when the messages we process contain them. It is for instance straightforwardly unproblematic to perceive the inconsistency of contradictory or paradoxical sets of information (e.g., there is no particular difficulty in seeing the problem with the following formulation “Buenos Aires is the capital of Argentina and Buenos Aires is not the capital of Argentina”), to reject information that is blatantly false (e.g. “ $2 + 2 = 22$ ”), or to recognise that certain conclusions do not follow from certain premises (e.g. “All Dutchmen are Europeans. Laszlo is European. Therefore, Laszlo is Dutch”).

Verbally communicated information unfortunately does not always come in such “simple” forms. Complex argumentative operations, vague formulations, different sorts of propositions requiring sophisticated (meta)representations (about states of affairs, desirable states of affairs,

8 Recently, Pilgram (2012) distinguished (in a pragma-dialectical analysis) between an argument *by* authority and *from* authority. An argument *by* authority is an argument in which the arguer refers to her own authority to support a (sub)standpoint (such as when a doctor refers to her long clinical expertise in arguing for a given treatment). An argument *from* authority consists in an appeal to an expert external to the present discussion in support of a (sub)standpoint – a case we refer to.

mental states, etc.), requirements of field-specific background knowledge and further types of constraints might well complicate the task of being epistemically vigilant. Just because we are able to spot blatant inconsistencies quite effortlessly does not mean that we are unfailingly good at spotting any inconsistency. Among other parameters affecting consistency assessment, motivation, attention and relevance might act as powerful constraints.

Mercier & Sperber (2009, 2011) have devoted specific attention to one particular aspect of epistemic vigilance, namely argument-processing. Their argumentative theory of reasoning posits the existence of a dedicated argumentative module whose role is to take a claim and information relevant to its evaluation as input and to yield a judgment about the relationship between premises and conclusion, i.e. a judgment about the argument's validity or acceptability. When it is first solicited, the argumentative module delivers an *intuition* about the argument's validity. This is an output we are not typically aware of: at the end of the process, we accept or reject the claim without necessarily being aware of the reasons we have for doing so. However, this does not mean that these reasons are not there; they simply exist without us being aware of them. Nevertheless, when required, the argumentative module can disembed the conclusion of the argument to keep on processing it reflectively, thereby consciously attending to the reasons why the argument should be accepted in the first place. This is reasoning proper.<sup>9</sup>

This line of research is evidently of great interest to argumentation theory; however, what we would like to focus on here, while intimately related to the argumentative module, does not directly concern its internal machinery. We will instead be concerned with what the argumentative module takes as input. To the extent that assessing message consistency requires processing verbal information, it is reasonable to assume that “comprehension of the content communicated by an utterance is a precondition for its acceptance” (Sperber *et al.* 2010: 367). In any argumentative exchange, recipients indeed need, at least minimally, to carry out two cognitive tasks: they need to understand the content of both the premises and the conclusion (i.e., they need to understand the meaning of the speaker's complete argumentation) and they need to be able to produce some sort of evaluation – even if it remains

9 See Mercier & Sperber (2009, 2011) and Mercier (2012) for an in-depth discussion.

intuitive at first – about whether the conclusion follows from its premises. The idea we want to develop here, with the straw man fallacy as a case in point (see sections 3 and 4), is that mechanisms of comprehension can also be constrained, thereby affecting the subsequent validity judgment yielded by the argumentative module. In particular, we contend that many fallacies do not operate on the justificatory link *per se* but rather on the (non)selection of critical information. Following Maillat & Oswald's (2009, 2011) model of the Contextual Selection Constraint (henceforth CSC), we will assume that language users who are processing arguments in real time and on the spot may be misled into selecting information sets that are irrelevant to the evaluation of an argument – and in some cases even detrimental to the latter.

The CSC holds that in principle the meaning derivation procedure can be constrained along two dimensions corresponding to Sperber & Wilson's extent conditions of relevance (1995: 125), namely processing effort and cognitive effect. Processing effort translates into the amount of resources it takes to represent a given assumption and is consequently determined by information accessibility: the more accessible the assumption, the less effort it will take to process, and the more chances it has of being kept in the meaning derivation procedure. Cognitive effects are defined in epistemic terms by Sperber and Wilson: the assumptions that will provide the most reliable information about the world (in terms of addition of new information, or revision and suppression of old information) will be the most relevant to the cognitive system. Those assumptions that achieve the best ratio between effort and effect are deemed the most relevant. According to Relevance Theory, maximisation of relevance so conceived is a general feature of cognition; as a consequence, any selection of information (be it explicit information contained in a verbal stimulus or contextual information an addressee is responsible to mobilise on his own) will be governed by this mechanism.

Accordingly, discursive strategies that can be exploited to manage the effectiveness of fallacious arguments can either *strengthen* (that is, make more accessible in terms of processing effort or epistemically stronger in terms of cognitive effects) information sets which are devoid of critical information or *weaken* (that is, make less accessible or epistemically weaker) information sets which are loaded with critical information and which, if represented, would tentatively defeat the fallacious move by attracting attention on its

fallaciousness.<sup>10</sup> The key point here is that these are constraints operating on *meaning*, to the extent that they manipulate both the accessibility and the strength of the information sets required to understand the message, which are the two extent conditions for reaching a relevant interpretation of a message. In view of that, we will show that the output of the comprehension procedure can be constrained and that this might have crucial implications for the overall judgment of argument validity or fallaciousness the recipient will be led to formulate (see the example in section 4 for a discussion).

### 2.2.3 *Cognitive heuristics*

A final note before we move on to the case of the straw man fallacy is in order. Our description of argument processing in light of recent work in cognitive psychology and on the study of verbal deception so far has highlighted how the way we process information can be constrained so as to influence the persuasiveness of verbal messages. How does that relate to cognitive heuristics?

Heuristics are traditionally defined as rules of thumb that help us take decisions, make judgments, and that more generally assist us in dealing with information in a cost-effective way. They are fast and frugal cognitive procedures that usually yield satisfactory results without requiring excessive amounts of cognitive resources:

Heuristics are frugal – that is, they ignore part of the information. Unlike statistical optimization procedures, heuristics do not try to optimize (i.e., find the best solution), but rather satisfy (i.e., find a good-enough solution). Calculating the maximum of a function is a form of optimizing; choosing the first option that exceeds an aspiration level is a form of satisficing. (Gigerenzer 2008: 20)

Traditionally, heuristics are associated with intuitive cognitive procedures, as opposed to effortful reflective ones. In this sense, both the comprehension

10 We should also mention here that the exploitation of these parameters is not an exclusive property of fallacies: sound arguments (and all discourse, actually), to the extent that they are comprehended following the same processing mechanisms that fallacious ones, are also subjected to these parameters. The difference between both is that fallacies block access to critical resources, while sound arguments in principle should foster it. The specificity of fallacies is thus to be found in their ability to obfuscate critical information.



module and the argumentative module may be considered to involve heuristics, to some extent at least.

We cannot help understanding an audible message that was uttered in our mother tongue (unless we block our ears or unless our attention is completely and consciously focused on something else); our interpretation of speaker meaning is most of the time unconscious and above all, we usually intuitively take what we understood from a verbal message to be reliable – in the sense that we do not engage in systematic doubt about what we have understood. Moreover, comprehension is fallible, to the extent that misunderstandings may occur. These usually result from a mismatch between linguistically decoded and/or contextual information selected by the addressee and linguistically encoded and/or contextual information selected by the speaker as she was planning and formulating her utterance. As fallibility is among the core features of heuristics, comprehension can be described as a heuristic. The argumentative module is also of a heuristic nature because it yields an intuitive inference, which is also fallible. Sometimes we are convinced by ill-evidenced claims, and further (reflective) examination allows us to identify the flaw. The effectiveness of fallacious arguments can furthermore be taken as evidence for the heuristic nature of default argument processing.

In this section we have tried to characterise more precisely what it means for a fallacy to tap on cognitive heuristics. We have focused in particular on how fallacies can go past epistemic vigilance filters, notably by trying to display expertise and consistency. We interpreted this in terms of strengthening and weakening strategies, and we now turn to discuss these mechanisms through our account of the straw man fallacy, which we then illustrate by an example in section 4.

### 3. Accounting for the straw man fallacy

#### 3.1 *Walton's account*

In his paper on heuristics, Walton lists the straw man fallacy among twelve informal fallacies that “need to be analysed with defeasible argumentation

schemes” (2010: 174-175). The straw man is thus an abuse of the “argument from commitment” as defined by Walton, Reed and Macagno (2008: 335). In a contemporary paper, however, Walton takes a rather different, if not the opposite, stance:

With many of the informal fallacies, the problem of fallacy identification, analysis, and evaluation is made easier by the fact that the fallacy is closely related to a known argumentation scheme [...]. This resource is not available, however, in the cases of wrenching from context and straw man. These fallacies are more purely dialectical in nature. They basically arise from or result in the misrepresentation of one party’s position by another party with whom the first party is engaged in a dialogue. (Walton & Macagno 2010: 303-304)

We agree with the latter formulation. In our view (Lewiński 2011, Lewiński & Oswald 2013), the straw man cannot be a violation of an argument scheme, because strictly speaking the straw man is not an argument in the first place. Rather, it is an attack on someone else’s argument; this is why, for instance, a real question (not a rhetorical one) can convey a straw man but not an argument. In terms of the pragma-dialectical critical discussion, a straw man is a (fallacious) speech act performed by the critical antagonist, rather than by the argument-making protagonist. In our analysis this accounts for the “dialectical nature” of the straw man fallacy, as Walton and Macagno (2010) have put it. One way of dealing with Walton’s inconsistency in approaching the straw man is to treat the second formulation (Walton & Macagno 2010) as a misrepresentation of dialectical commitments in terms of the form of the fallacy, and the first formulation (Walton 2010) as a type of attack from commitment in terms of the function of the straw man. Only a clear grasp of both the form and the function of the straw man would define the fallacy, as argued by Lewiński (2011: 480). Walton, thus, would propose two mutually exclusive accounts of the straw man, since he actually speaks of two different aspects of the straw man that eventually can be seen as complementary, rather than exclusive. Such an interpretation, while highly charitable, is not explicitly put forth by Walton himself, and thus may not accurately capture his take on the fallacy. Noticeably, however, Walton (and Macagno) offer no explanation, in terms of argumentation schemes or *paraschemes*, of how straw man attacks are deceptive and may pass for good, reasonable forms of argumentative criticism.

In the following, we will propose a different take on the straw man fallacy that clearly relates fallacy evaluation to precisely described cognitive mechanisms. Our contribution is thus novel not in its general aim (since both Walton and Jackson have undertaken a similar line of inquiry), but rather in the exact way we envisage a fruitful cross-fertilisation of recent developments in normative theories of argumentation and cognitive studies. From the perspective of our goals, even if Walton and Jackson both link fallacies and heuristics, they are not attentive enough to cognitive mechanisms underlying the deceptiveness of many fallacies. Moreover, they do not offer an applicable normative account of fallacies that supports a consistent analysis of the straw man: while Jackson chooses a decidedly descriptive path, Walton seems to offer two incompatible accounts of how to best analyse a straw man as a fallacy. The advantage of our work is thus, first, a higher level of precision in both its normative and cognitive aspects, and, second, a focus on the detailed working of one particular fallacy – the straw man – that seems somewhat elusive in other accounts.

### 3.2 *An integrated normative and cognitive pragmatic account of the straw man fallacy*

In terms of the integrated pragma-dialectical theory, a straw man, similarly to any other fallacious argumentative move, is conceptualised as a “derailment of strategic manoeuvring” (van Eemeren & Houtlosser 2003). *Strategic manoeuvring* is defined as a continuous balancing between two competing yet reconcilable goals in argumentation. On the one hand, in order to resolve a difference of opinion in the process of genuine critical testing of their opinions, arguers should meet dialectical requirements of reasonableness embodied in the rules of the PD model of a critical discussion (van Eemeren & Grootendorst 2004: Chs. 6 and 8). On the other hand, in attempts to have their position accepted by the antagonist, arguers may take advantage of numerous rhetorical techniques. While manoeuvring aimed at reconciling the reasonable with the opportune in argumentation is perfectly possible, the strife for persuasive success may also lead to the abandonment of standards of reasonableness. Whenever the latter happens, pragma-dialecticians speak of a derailment of strategic manoeuvring that by definition amounts to

committing a fallacy, because a rule for a reasonable critical discussion is violated.

The basic PD understanding of the straw man fallacy is thus that of a violation of rule 3 for a critical discussion: *A party's attack on a standpoint must relate to the standpoint that has indeed been advanced by the other party* (van Eemeren & Grootendorst 1992: 124-131).<sup>11</sup> Contravening this rule and hence attacking a straw man seriously hinders critical testing. The antagonist who surreptitiously misrepresents the protagonist's opinions seems to be involved in genuine critical testing of these opinions, but in fact attempts no more than a made-up falsification aimed at rhetorical victory (traditionally termed a "sophistical refutation"). As a result, the possibility of reaching a reasonable resolution of the entire dispute on the merits is seriously undermined.

Yet, even though the possible rhetorical advantage that can be gained by committing fallacies is seen in PD as a driving force behind much of fallacious argumentative discourse, this theory does not aim at expounding how such fallacious persuasive appeals actually work.

We have consequently tried to establish elsewhere (Lewiński & Oswald 2013) the grounds for the complementarity of cognitive pragmatics (and Relevance Theory specifically) and PD in an exhaustive treatment of the straw man fallacy. We hypothesised that accounting for the straw man fallacy is a matter of interpretation of argumentative discourse *on two distinct levels*. The first level of interpretation, which is the normative level of argumentation analysis, is concerned with providing identifying criteria for the straw man fallacy. The second level of interpretation, dedicated to explain why straw men may be effective, considers mechanisms of comprehension and how these constrain what recipients of the straw man fallacy understand from the message. Specifically, we claim that a cognitive pragmatic model of utterance interpretation is ideally equipped to explain how

11 We note here that the rules of the critical discussion, as identified in PD, are not comparable to conversational maxims (cg. Grice 1989), whose flouting is instrumental to the derivation of meaning (in terms of implicature identification): pragma-dialectical rules ensure critically reasonable argumentative discussions and the recognition that these have been violated in a given discourse allows the analyst to pass a judgement on the fallaciousness of the argument. In other words, the rules of the critical discussion do not have anything to do with meaning.

misrepresentations involved in a straw man attack may pass for genuine and faithful representations.

We thus first stipulated certain general contextual conditions under which an analyst can judge a given interpretation of the protagonist's position a straw man. We formulated the first basic criterion of identification for the straw man fallacy in terms of pragmatic plausibility: as long as the antagonist follows certain contextual possibilities in deriving speaker meaning and thus as long as she stays within the bounds of a disagreement space of a given utterance, she cannot be seen as committing a straw man.<sup>12</sup> The second crucial criterion for normative interpretation is interpretative charity. Charity of interpretation – i.e., the choice of an interpretation that is most beneficial to the arguer – is typically advised to arguers and analysts alike in cases of interpretative doubt (see Lewiński 2011, 2012). However, rather than perceiving it as a rule of reasonable argumentation, we take charity between arguers to be a rhetorical choice which can be made one way or another. The antagonist who is capable of producing compelling criticisms against the protagonist's position may opt for a very charitable interpretation of the protagonist's argumentation. Yet, she can also opt for an uncharitable interpretation – and in principle there is nothing wrong with such an interpretation as long as it is pragmatically plausible. All the same, we contend that similarly to pragmatic plausibility the criterion of interpretative charity between arguers is inherently contextual. In some contexts that can be termed *constructive*, arguers are expected, indeed required, to be more charitable than in *critical* contexts. Similarly, some contexts call for heightened *precision* (thus limiting the scope of plausible interpretations), while others allow for *loose* interpretations where plausibility is much extended (see Lewiński 2011: 490ff.).

Even if satisfactorily laid out, however, such conditions pertain to an analyst's "normative interpretation" and evaluation based on a specific theory of argumentation. Therefore, such inquiry does not answer the question of how ordinary addressees, in the process of pre-theoretical, natural comprehension (or "naïve interpretation") are misled into

12 The notion of disagreement space denotes the "entire complex of reconstructible commitments" and consequently defines "a structured set of opportunities for argument" (van Eemeren *et al* 1993: 95). Pragmatically speaking, the disagreement space consists of all contextually plausible interpretations of a given utterance.

interpreting argumentative discourse in a way that will leave the straw man unidentified. Building on a cognitive pragmatic model of information-selection, we argued along the lines exposed above in section 2.2.2 that straw men fallacies effectively manage to constrain meaning so that misattributions of commitment pass for correct attributions of commitment. This process can unfold along the abovementioned two types of constraining strategies: the relevance of potential clues revealing the misattribution of commitment can be weakened, while evidence of a sound attribution of commitment is strengthened.

Building on the complementarity of the two pragmatic approaches used, we gave a fuller account of the working of the straw man fallacy, in terms of both its dialectical incorrectness and rhetorical treacherousness; in what follows we turn to discussing a real example to illustrate our account.

#### 4. Exhuming the straw man: a concrete example

During one of the meetings of the Swiss National Council in December 2008, Carlo Sommaruga, a representative for the canton of Geneva, officially submits and motivates a motion meant to legally rehabilitate the Swiss fighters of the International Brigades, who fought in Spain against Franco's dictatorship in the 1930s.<sup>13</sup> Most of these fighters, upon their return to Switzerland, were prosecuted on the grounds that the Swiss military criminal code does not authorise its citizens to serve in foreign armies, and most of them were sentenced to prison (with sentences ranging from 15 days to 4 years of imprisonment, together with a denial of their civic rights). Oskar Freysinger, another member of the lower house of the Swiss parliament, opposes this motion with the following argumentation:

13 The official text of the motion (in French) can be found online in the archives of the Swiss parliament: [http://www.parlament.ch/ab/frameset/f/n/4806/284915/f\\_n\\_4806\\_284915\\_285114.htm](http://www.parlament.ch/ab/frameset/f/n/4806/284915/f_n_4806_284915_285114.htm) (Last accessed 01.05.2013).

Mr. Sommaruga, you are painting a very romantic portrait of the International Brigades. However, did you know that George Orwell, the very well-known author who fought in their ranks, very quickly noted that these people aimed at the establishment of a Stalinist regime and that their violence was as serious as the one perpetrated by the fascists? So, do you want to rehabilitate people who intended to fight a totalitarian regime in order to replace it with another totalitarian regime?<sup>14</sup>

The last sentence of Freysinger's intervention, which is realised as a rhetorical question, can be taken to convey both the standpoint of his argumentation and a simple justification for it; we give three reasons in support of this reading. First, in contemporary democratic countries such as Switzerland, no reasonable person would indeed support a totalitarian regime, based on the widespread idea that totalitarianisms are undesirable political systems. Second, the information provided before the question and which relates to the International Brigades' violence is presented by Freysinger as evidence in support of their non-rehabilitation. Finally, Freysinger's contribution is immediately consecutive to Sommaruga's exposition of the motion: in the broader context of this particular activity type, Freysinger's question functions as an opposition to what was proposed.<sup>15</sup> For these reasons, we take the main standpoint of the argument to be the following:<sup>16</sup>

1. Standpoint: The Swiss fighters of the International Brigades should not be legally rehabilitated.

Several arguments are presented in support of the claim. Looking at the rhetorical question, it appears that according to the speaker these fighters were seeking to replace a totalitarian regime by another totalitarian regime,

- 14 Here is the original text in French: "*Monsieur Sommaruga, vous nous dressez un portrait très romantique des Brigades internationales. Or, savez-vous que George Orwell, auteur très connu, qui a combattu dans leurs rangs, a très vite remarqué que ces gens visaient à l'établissement d'un régime stalinien et que leurs violences étaient tout aussi graves que celles que commettaient les fascistes? Alors voulez-vous réhabiliter des gens qui cherchaient à combattre un régime totalitaire afin de le remplacer par un autre totalitarisme?*".
- 15 Note that exactly because Freysinger's standpoint is the contradictory of Sommaruga's proposal, the difference of opinion is "mixed" (see van Eemeren & Grootendorst 2004: 119-120): Freysinger builds his argumentation to both support his position and criticise Sommaruga. This situation warrants an analysis of his arguments in terms of the straw man, as defined above.
- 16 We are following the PD notation of the structure of arguments in an analytic overview. See van Eemeren *et al.* (1993).

and this, combined with an implicit premise pointing to the undesirability of this state of affairs, gives us a first argument in support of (1):

- 1.1a      Explicit premise: The International Brigades wanted to replace a totalitarian regime by another totalitarian regime.
- (1.1a')    Implicit premise: We should not legally rehabilitate people who want to replace a totalitarian regime by another totalitarian regime.

What precedes the rhetorical question functions as evidence for its content: Freysinger is using what he knows about the International Brigades to legitimise the crucial idea that the fighters of these brigades were promoters of a totalitarian regime. This means that the explicit premise (1.1a) becomes in our reconstruction of the argumentative chain the standpoint of a sub-argument. In what precedes, we do find at least one relevant required premise, which, combined with an additional implicit premise accessed through our encyclopaedic knowledge, allows us to reach (1.1a); to the extent that the conclusion from (1.1a.1) and (1.1a.1') is not exactly worded as (1.1a), we represent it in (1.1a\*) as an intermediate step allowing us to spell out the link between (1.1a.1) and (1.1a.1') on the one hand and (1.1a) on the other:

- 1.1a.1     Explicit premise: These people aimed at the establishment of a Stalinist regime.
- (1.1a.1')    Implicit premise: The Stalinist regime was totalitarian.
- 1.1a\*      Implicit standpoint: These people aimed at the establishment of a totalitarian regime.

This standpoint is not exactly formulated as the premise we identified in (1.1a); however, both (1.1a\*) and (1.1a) have in common the idea that the International Brigades, according to Freysinger, wanted to bring a totalitarian regime to power. Moreover, they are both supported by the same premises that taken together satisfy the “logical minimum” (van Eemeren & Grootendorst 2004: 117-118) of argumentation. Next to this sub-argument, Freysinger also notes that the violent nature of the International Brigades is as serious as the violent actions committed by the fascists. This information becomes relevant by acquiring an argumentative function, though this time not on a subordinate level, but directly with respect to the main standpoint (1), as it contributes an additional reason not to legally rehabilitate the Swiss fighters. Here is the reconstruction of this particular argument with an



explicit formulation of the implicit premise required to support the standpoint:

- 1.1b      Explicit premise: The violence of the International Brigades was as serious as the one perpetrated by the fascists.
- (1.1b')    Implicit premise: We should not legally rehabilitate those whose violent actions were as violent as the actions of the fascists.
- 1          Standpoint: The Swiss fighters of the International Brigades should not be legally rehabilitated.

Breaking the argumentation further down into its different parts, we notice that both (1.1a.1) and (1.1b) gain their epistemic strength from their contents being attributed to George Orwell, “the very well-known author”. Here, Freysinger is trying to make the two premises epistemically stronger by taking advantage of Orwell’s credit in its quality of direct witness of what was happening in the ranks of the International Brigades – after all, he was one of them and he was there, as asserted by Freysinger. It is not our purpose here to assess in detail whether this appeal to authority is fallacious or not, even though we think there is considerable room to debate the issue.<sup>17</sup> It remains that the use of Orwell’s authority by the speaker intervenes as a way of establishing the acceptability of premise (1.1a.1) and premise (1.1b), which are part of two coordinated argumentations in support of the main standpoint (1).

A careful reconstruction and synthesis of the argument leads us to the following global paraphrase:

- 2.          We should not legally rehabilitate the Swiss fighters of the International Brigades because they wanted to replace a totalitarian regime by another totalitarian regime (since, as noted by Orwell, they promoted a Stalinist

17 We have no conclusive evidence that the appeal to authority here is fallacious, but we note that Orwell’s credibility as an author does not give him the authority to judge if the International Brigades were violent or if they wanted to establish a totalitarian regime; furthermore, his taking part to the action of the International Brigades is not sufficient either to assess his credibility – we would need to know when he was part of them, his motivation for joining them, and his reasons for leaving them. Still, an *ad verecundiam* argument seems to be lurking here and it is probably intended by Freysinger to make (1.1a.1) and (1.1b) epistemically stronger than other assumptions. See also Herman (this volume) for a detailed discussion of arguments from authority.

regime) and because – and this is also noted by Orwell – their violence was as serious as the violence perpetrated by the fascists.

There is still one step missing in our analysis, and it has to do with the link that Freysinger creates between the International Brigades wanting to promote totalitarianism and the fact that they were promoting Stalinism. In other words, we are concerned here with the shift from premises (1.1a.1) and (1.1a.1') to the conclusion (1.1a\*). While (1.1a.1) is, as we have seen, supported by the appeal to Orwell's authority, nothing is adduced in support of the implicit premise (1.1a.1'), which is nevertheless crucially needed to reach the conclusion (1.1a\*), itself subsequently needed to reach the main standpoint of the argument (1). And this is precisely where we contend that the straw man is hidden. Unexpressed premise (1.1a.1') tells us that Stalinism was a totalitarian regime. Indeed, nowadays we know that such was the case, and that the concentration of power into one man's hands led the former Soviet Union to suffer the worst dictatorship of its history. Yet, is it fair to consider that the Swiss fighters of the International Brigades were supporting a totalitarian regime? Freysinger tells us that they supported Stalinism, which is true. But in the 1930s context, did supporting Stalinism straightforwardly amount to supporting a dictatorship? This is far from certain. It is only after the 1930s that the world gradually started to find out about the horrors of Stalinism; before that, few had witnessed them, and it would probably be more accurate to say that the International Brigades had as an ideal the promotion of a regime which, on paper, was the complete opposite of a dictatorship. Furthermore, in the 1930s, communism under Stalin had not yet become a totalitarian regime – so there was no way people could know it. Therefore, accusing the Swiss fighters of promoting a totalitarian regime is uncharitable – and furthermore anachronistic since they could not know that Stalinism was going to become such a regime. Had they known what Stalinism became, they would most likely never have supported it, since that is precisely the type of regime they were fighting at that time (i.e., Franco's fascist regime).

In this example, Freysinger is using a problematic piece of encyclopaedic knowledge in (1.1a.1') as an implicit premise to support the implicit conclusion (1.1a\*), which is itself, as a misrepresentation of the Swiss fighters' motivations, used in the global argument in order to attack Sommaruga's main standpoint. The straw man in this example seems to be

multi-layered and complexly structured: its first step consists in misrepresenting the intentions of a group of people targeted by the motion, namely the fighters of the International Brigades; we have seen that the misrepresentation can be qualified as such because it is both uncharitable, but also pragmatically implausible by virtue of its anachronistic nature. It should be highlighted in this respect that the treacherousness of the argumentative move here does not lie in the explicit premise (1.1a.1), because the International Brigades indeed wanted a Stalinist regime. The problem is that in the 1930s context (1.1a.1') was not known, and therefore it should not legitimately be counted among something these fighters could reasonably be taken to believe.

But the actual straw man in the parliamentary discussion in question, instantiated here as an attempt to counter Sommaruga's standpoint, only becomes fully fledged as Freysinger formulates his rhetorical question. Through this question, Freysinger implies that Sommaruga endorses the (misattributed) views of the International Brigades about promoting Stalinism. The move consequently also results in a misattribution of commitment to Sommaruga: as Freysinger puts it, his opponent is guilty of intending to rehabilitate people who were no better than fascists. As noted above, such a criticism is not only uncharitable – which is expected of two deputies from opposing parties and thus contextually acceptable – but also implausible due to an anachronistic ascription of intentions: one could hardly be a supporter of totalitarian Stalinist methods before these methods became known or even existed. Therefore, Freysinger's critique of Sommaruga's motion does qualify as a straw man fallacy, even though of a rather subtle, pragmatic type. To sum up on the subtlety – and complexity – of the straw man in Freysinger's intervention: his misattribution of beliefs to the fighters of the International Brigades is not a straw man fallacy *per se*, to the extent that there is no argumentative discussion between him and these fighters. However, by misrepresenting their beliefs and by implicitly communicating that Sommaruga is liable to endorsing them through his motion, Freysinger makes as if these misrepresentations are also part of Sommaruga's actual commitments.

A closer look at the linguistic choices used by Freysinger in this complex argumentation will reveal how the relevance of critical information is weakened, while the fallacy's apparent relevance as a sound argument is at

the same time strengthened. First of all, the rhetorical question is directly addressed at Sommaruga in a vocative mode (“Do *you* want to rehabilitate...”). It would only be relevant to ask such a question if there were plausible reasons to consider that Sommaruga indeed had voiced the intention of rehabilitating people who promoted a totalitarian regime. Moreover, a rhetorical question, as pointed out by Snoeck Henkemans, “serves as a means to urge the addressee to act on his commitment and recognise that the standpoint that is being defended by the argument the addressee supposedly accepts, should now also be accepted” (Snoeck Henkemans 2009: 17); here Freysinger is (deceptively) trying to get Sommaruga to endorse the implications of one of his alleged previous commitments for argumentative purposes. The question under discussion is hence not at all whether the International Brigades promoted a totalitarian regime, but rather whether Sommaruga is prepared to accept that such was the case. By constraining the question’s focus onto Sommaruga’s commitments, Freysinger may be charged of attempting to direct his audience’s attention (including his opponent’s) on the attack rather than on the grounds of the attack. In other words, the rhetorical question *strengthens* the relevance of Sommaruga’s supposed mistake, because, contextually, that’s the very point of Freysinger’s intervention. As for the grounds on which the attack is launched, their relevance in the exchange is made less salient and is *ipso facto* weakened.

From the perspective of strengthening, an additional parameter appears to be of importance: the *ad verecundiam* we identified earlier is meant to strengthen the epistemic status of (1.1a.1) as it attempts to take advantage of Orwell’s presumed credibility. By resorting to this strategy, Freysinger intends to increase the chances of (1.1a.1) being accepted as a reliable piece of information. Then, once (1.1a.1) is accepted and after (1.1a\*) is processed as an explicit content of the rhetorical question, the only way to understand the argumentative link, or, in other words, the relevance of (1.1a.1) with respect to (1.1a\*) is to represent the implicit premise (1.1a.1’). In other words, in order to *understand* the link between (1.1a.1) and (1.1a\*), we need to mobilise (1.1a.1’) – albeit only intuitively. And since (1.1a.1’) is necessary for comprehension purposes, it might be argued that it will be part of the meaning derivation procedure, thereby increasing its chances of being selected. Crucially, the use the audience is asked to (intuitively) make of

(1.1a.1') is not related to any kind of argumentative evaluation; it is instead presented as a condition for the meaning of Freysinger's contribution to be properly understood. We claim that this weakens the chances of (1.1a.1') getting "argumentatively" picked up. In other words, the addressee's cognitive environment is encouraged to consider that (1.1a.1') is relevant prominently with respect to the comprehension procedure, and not so much with respect to the argumentative procedure.<sup>18</sup>

More can be said on the particular status of (1.1a.1'). The second parameter affecting the salience of this first commitment misattribution is its locus of occurrence in the overall argumentation. (1.1a.1') is part of a subordinate argument which is used in support of one of the coordinated arguments (1.1a) adduced in support of the main standpoint in (1). The position of this misattribution is probably not fortuitous; since it is embedded in a sub-argument – and moreover not presented as a piece of information that needs to be called into question due to its prominent interpretative role –, it is not presented as decisive for the global argument. Again, this can be interpreted in terms of salience: for the whole argument to make sense, (1.1a.1') need not be consciously represented, which reduces the chances of the content being submitted to critical doubt. Put differently, (1.1a.1') is just a mere background to the main argumentation. As such, it is presented as if its weight was not pivotal in the overall strength of the case.

Summing up, we contend that the example of Freysinger's criticism is an interesting case of the straw man fallacy. It remains rather subtle and thus requires close critical attention within a framework of a context-sensitive,

18 The nature of (1.1a.1') in terms of meaning components can be further discussed. While it seems pretty clear, as argued by an anonymous reviewer, that (1.1a.1') qualifies as an implicated premise, there could be room to suspect that it is a *discursive* presupposition (distinct from classical semantic presuppositions, see Saussure, 2012, forth. and this volume for a discussion of the notion): its function, in addition to its argumentative role, would be to provide the necessary bits of information for the link between (1.1a.1) and (1.1a\*) to plainly make sense. Interestingly, the contribution of presuppositions to meaning is subtle and occurs somewhat "backstage". Some have claimed that the contents of presupposition can accordingly be forced into communication while escaping the addressee's critical awareness (see Saussure 2012, and Polyzou 2013 for instance). We also take this to be an indication that the tentative discursive presuppositional status of (1.1a.1') might be processed by the comprehension module and fail to qualify as input for the argumentative module.

normative pragmatic theory, such as PD, to be clearly assessed as fallacious. But because of this subtlety, the example also vividly illustrates the different weakening and strengthening strategies employed in Freysinger's discourse that may affect the way the argument is processed by the audience of his attack. A first misattribution of commitment, whose argumentative potential is obscured by its importance in the meaning derivation procedure, is used to establish the grounds on which the primary straw man fallacy operates. In other words, comprehension has constrained argumentation. By implicitly leaving room for the audience to infer that Sommaruga's motion is an indication that he might endorse totalitarian views himself, Freysinger tries to push his opponent into defending himself, which might turn into a "cognitive priority", at the expense of a critical examination of the argument.

## 5. Conclusion

Accounting for the straw man fallacy is a clear example of the need for pragmatically-driven theoretical research. In this paper we have tried to show how a combination of two contemporary theories with strong pragmatic import may provide a grounded alternative to existing accounts by addressing both normative and cognitive questions.

Reliance on PD makes for the requirements of the question of fallacy identification; reliance on cognitive pragmatics but also on cognitive psychology and anthropology makes for the requirements of the question of fallacy effectiveness. It was one of the purposes of this paper to highlight the role that cognitive studies should be able to endorse within the framework of argumentation theory.

Underlying our exposition of the persuasive effectiveness of the straw man fallacy are questions of great cognitive significance; notably, we touched here upon the complex issue of the way verbal stimuli are processed, depending on their relevance for different cognitive tasks. The example discussed in section 4 has shown that the same representation (in this case the implicit premise (1.1a.1')) might be processed either by the comprehension module or by the argumentative module. Crucially, we

argued that in the example the interpretative path needed to be privileged in order for the fallacy not to be critically tested, that is, in order to block its processing by the argumentative module. It might be an indication that variations in communicative contexts might trigger competition between cognitive modules. This could well represent a new direction of research in the study of rhetorical effectiveness.

## References

- Aikin, S. J. & Casey, J. (2011): "Straw men, weak men, and hollow men", *Argumentation* 25 (1), 87-105.
- Bardone, E. (2011): *Seeking chances. From biased rationality to distributed cognition*. Springer, Berlin/Heidelberg.
- Bizer, G., Kozak, S. & Holterman, L. (2009): "The persuasiveness of the straw man rhetorical technique", *Social Influence* 4 (3), 216-230.
- Burge, T. (1993): "Content preservation", *Philosophical Review* 101, 457-88.
- Correia, V. (2011): "Biases and fallacies: The role of motivated irrationality in fallacious reasoning", *Cogency* 3 (1), 107-126.
- Eemeren, F. van & Grootendorst, R. (2004): *A Systematic Theory of Argumentation: The Pragm-dialectical Approach*. Cambridge University Press, Cambridge.
- (1992): *Argumentation, Communication, and Fallacies: A Pragm-dialectical Perspective*. Lawrence Erlbaum, Hillsdale, NJ.
- Eemeren, F. van, Grootendorst, R., Jackson, S. & Jacobs, S. (1993): *Reconstructing argumentative discourse*. The University of Alabama Press, Tuscaloosa/London.
- Eemeren, F. van & Houtlosser, P. (2003): "Fallacies as derailments of strategic maneuvering: The *argumentum ad verecundiam*, a case in point", in Eemeren, F.H. van, Blair, J.-A., Willard, C.A., Snoeck Henkemans, F. (eds), *Proceedings of the Fifth Conference of the International Society for the Study of Argumentation*. SicSat, Amsterdam, 289-292.
- Evans, J.S.B.T. (2004): "Biases in deductive reasoning", in Pohl, R. (ed.) *Cognitive illusions. A handbook on fallacies and biases in thinking, judgement and memory*. Psychology Press, Hove/New York, 127-144.
- Evans, J.S.B.T. & Frankish, K. (eds) (2009): *In Two Minds*. Oxford University Press, New York.
- Gigerenzer, G. (2008): "Why Heuristics Work", *Perspectives on psychological science* 3 (1), 20-29.
- Gigerenzer, G., Hertwig, R. & Pachur, T. (eds) (2011): *Heuristics: The foundation of adaptive behavior*. Oxford University Press, New York.
- Grice, H. P. (1989): *Studies in the way of words*. Harvard University Press, Cambridge, MA.
- Hamblin, C. (1970). *Fallacies*. Methuen, London.

- Herman, T. (2014). "L'argument d'autorité: structure, définition et effets rhétoriques", in Herman, T. & Oswald, S. (éds), *Rhétorique et cognition*. Peter Lang, Berne.
- Jackson, S. (1996): "Fallacies and heuristics", in Bentham, J. van, Eemeren, F.H. van, Grootendorst, R. & Veltman, F. (eds), *Logic and argumentation*. Royal Netherlands Academy of Arts and Sciences, Amsterdam, 101-113.
- Kahneman, D. & Tversky, A. (1972): "Subjective probability: a judgement of representativeness", *Cognitive Psychology* 3, 430-454.
- Kahneman, D. & Tversky, A. (1973): "On the psychology of prediction", *Psychological Review* 80, 237-251.
- Laar, J. van (2008): "Room for maneuver when raising critical doubt", *Philosophy and Rhetoric* 41 (3), 195-211.
- Lewiński, M. (2011): "Towards a critique-friendly approach to the straw man fallacy evaluation", *Argumentation* 25 (2), 469-497.
- (2012): "The paradox of charity", *Informal Logic* 32 (4), 403-439.
- Lewiński, M. & Oswald, S. (2013): "When and how do we deal with straw men? A normative and cognitive pragmatic account", *Journal of Pragmatics* 59 (Part B), 164-177.
- Maillat, D. (2013): "Constraining context selection: on the pragmatic inevitability of manipulation", *Journal of Pragmatics* 59 (Part B), 190-199.
- Maillat, D. & Oswald, S. (2009): "Defining manipulative discourse: the pragmatics of cognitive illusions", *International Review of Pragmatics* 1 (2), 348-370.
- (2011): "Constraining context: a pragmatic account of cognitive manipulation", in Hart, C. (ed.), *Critical discourse studies in context and cognition*. John Benjamins, Amsterdam, 65-80.
- Mercier, H. (2012): "Looking for arguments", *Argumentation* 26 (3), 305-324.
- Mercier, H. & Sperber, D. (2009): "Intuitive and reflective inferences", in Evans, J.S.B.T. & Frankish, K. (eds), *In Two Minds*. Oxford University Press, New York, 149-170.
- (2011): "Why do humans reason? Arguments for an argumentative theory", *Behavioral and Brain Sciences* 34 (2), 57-74.
- Millikan, R. G. (1987): *Language, Thought and Other Categories*. MIT Press, Cambridge, MA.
- Oswald, S. (2011): "From interpretation to consent: Arguments, beliefs and meaning", *Discourse Studies* 13 (6), 806-814.
- Oswald, S. (forth.): "It is easy to miss something you are not looking for: A pragmatic account of covert communicative influence for (critical) discourse analysis", in Hart, C. & Cap, P. (eds), *Contemporary Studies in Critical Discourse Analysis*. Continuum, London.
- Pilgram, R. (2012): "Reasonableness of a doctor's argument by authority: A pragma-dialectical analysis of the specific soundness conditions", *Journal of Argumentation in Context* 1 (1), 33-50.
- Reid, Th. (2000) [1764]: *An Inquiry into the Human Mind and the Principles of Common Sense*. Penn State Press, University Park, PA.
- Saussure, L. de (2013): "Background relevance", *Journal of Pragmatics* 59 (Part B), 178-189.
- Saussure, L. de (2014). "Présuppositions discursives, assertion d'arrière-plan et persuasion", in Herman, T. & Oswald, S. (éds), *Rhétorique et cognition*. Peter Lang, Berne.
- Snoeck Henkemans, F. (2009): "Manoeuvring strategically with rhetorical questions", in van Eemeren, F.H. & Garssen, B. (eds), *Pondering on problems of argumentation*. Springer, New York, 15-23.



- Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G. & Wilson, D. (2010): “Epistemic vigilance”, *Mind and Language* 25 (4), 359-393.
- Sperber, D. & Wilson, D. (1995) [1986]: *Relevance. Communication and cognition*. Blackwell, Oxford.
- Talisse, R. & Aikin, S. F. (2006): “Two forms of the straw man”, *Argumentation* 20 (3), 345-352.
- Tversky, A. & Kahneman, D. (1974): “Judgment under uncertainty: Heuristics and biases”, *Science*, New Series 185 (4157), 1124-1131.
- Walton, D. (1996): “The straw man fallacy”, in Bentham, J. van, Eemeren, F.H. van, Grootendorst, R., Veltman, F. (eds), *Logic and Argumentation*. Royal Netherlands Academy of Arts and Sciences, Amsterdam, 115-128.
- Walton, D. (2010): “Why fallacies appear to be better arguments than they are”, *Informal Logic* 30 (2), 159-184.
- Walton, D., Reed, C. & Macagno, F. (2008): *Argumentation Schemes*. Cambridge University Press, New York.
- Walton, D. & Macagno, F. (2010): “Wrenching from context: The manipulation of commitments”, *Argumentation* 24 (3), 283-317.
- Wason, P. (1960): “On the failure to eliminate hypotheses in a conceptual task”, *The Quarterly Journal of Experimental Psychology* 12 (3), 129-140.
- (1966): “Reasoning”, in Foss, B. M. (ed.), *New horizons in psychology*. Penguin, Harmondsworth, 106-137.
- Woods, J. (2004): *The death of argument: Fallacies in agent based reasoning*. Kluwer Academic Publishers, Dordrecht.



## Résumés / Abstracts

### La vigilance épistémique

Dan SPERBER, Fabrice CLÉMENT, Christophe HEINTZ, Olivier MASCARO, Hugo MERCIER, Gloria ORIGGI et Deirdre WILSON

Les êtres humains dépendent massivement de la communication avec autrui, malgré le risque que cela comporte d'être accidentellement ou intentionnellement mal informé. Pour conserver le caractère avantageux de la communication, en dépit de ce risque, nous défendons l'idée que les êtres humains disposent d'une série de mécanismes cognitifs dédiés à la vigilance épistémique. Nous développons cette idée ici et illustrons comment cette vigilance épistémique fonctionne dans plusieurs aspects de la vie mentale et sociale en examinant des questions, recherches et théories dans différents domaines de la philosophie, de la linguistique, de la psychologie cognitive et des sciences sociales.

### Manipulation et cognition

Didier MAILLAT

Cette contribution propose un modèle cognitif pour une catégorie de phénomènes langagiers qui appartiennent au discours manipulateur (voir Saussure & Schulz 2005, van Dijk 2006, van Eemeren & Grootendorst 2004). Elle offre une nouvelle approche de la manipulation dans le cadre d'un modèle pragmatique cognitiviste inspiré des travaux en théorie de la pertinence (Carston 2002, Sperber & Wilson 1995). L'argument central consiste à traiter la manipulation comme une stratégie discursive visant à exploiter de façon systématique les limitations cognitives affectant les processus d'interprétation. Sperber, Girotto & Cara (1995) ont fait référence à ces limitations du système cognitif comme étant la résultante indésirable d'un 'optimisme cognitif'. La manipulation est envisagée comme une manœuvre visant à contrôler le processus de sélection de contexte décrit dans la théorie de la pertinence, de manière à empêcher le locuteur par un processus de contrainte cognitive à opérer une sélection optimale des éléments de son environnement cognitif. La manipulation est ainsi présentée

comme exploitant des ‘défauts’ inhérents aux processus cognitifs qui gouvernent l’interprétation (voir Maillat & Oswald 2009, 2011).

Dans ce cadre, le chapitre présente un modèle pragmatique des stratégies manipulatoires a) dont les hypothèses sont testables expérimentalement b) et qui dresse un parallèle systématique entre ces dernières et une catégorie de phénomènes qui a déjà été abondamment étudiée dans le domaine de la psychologie cognitive, à savoir les illusions cognitives (voir Pohl 2004).

Par le biais de ce rapprochement, une ré-analyse de motifs manipulatoires connus, tels que les arguments fallacieux (van Eemeren & Grootendorst 2004), peut être entreprise dans le cadre du modèle pragmatique proposé, laquelle ne se contente pas de décrire des phénomènes linguistiques de surface, mais propose de les expliquer au travers de certains mécanismes cognitifs sous-jacents. Dans cette dernière partie, je m’intéresserai donc plus spécifiquement aux bases cognitives d’arguments fallacieux (e.g. *ad verecundiam*, *ad populum*) et – inversement – aux échos dans le discours manipulateur de certaines illusions cognitives (e.g. *mere exposure effect*, *labelling effect*, *anchoring effect*).

## Biased argumentation and critical thinking

Vasco CORREIA

Assuming that logical and dialectical theories suffice to ensure the rationality of argumentation, the problem remains that arguers systematically transgress these rules in real-life debates. This may happen, to be sure, because arguers deliberately resort to illegitimate means of persuasion, but there is considerable evidence that fallacious reasoning is also rooted in a variety of cognitive illusions that tend to occur unintentionally (biases and heuristics). Insofar as these illusions typically occur without the arguer’s awareness, even those who try to follow the proper rules of reasoning may end up arguing in a biased manner. The person who “rationalizes”, for example, genuinely tries to justify her attitudes in a logical and reasonable fashion but typically commits the fallacy of misidentifying the cause. Furthermore, certain biases do not even entail a clear violation of argumentation rules, and yet lead to partial and unreasonable views. There is nothing illogical, for example, about focusing too much on evidence that confirms our own preconceptions (confirmation bias), but this gives rise to all sorts of prejudiced and blinkered views. Some of these biases are due to cognitive malfunctioning, while others

seem to be motivated by the arguer's emotions and interests. In either case, biased argumentation undermines the reasonableness of discussions and tends to aggravate the problem of polarization of opinions. I argue that if normative theories of argumentation are to be effective instruments with practical significance, rather than mere ideals of how people ought to argue, they need to take into account what empirical findings reveal regarding the type of errors that people actually commit in everyday life debates. More specifically, my hypothesis is that there are numerous correlations between specific types of fallacies and specific types of cognitive illusions, and that an understanding of the mechanisms underlying such illusions is key to preventing irrational thinking in argumentative contexts. Although cognitive and motivational biases typically occur at a sub-intentional level, it remains possible to counteract such biases indirectly, through the adoption of second-order strategies of argumentative self-control.

### Vers une naturalisation de la rhétorique ? Problèmes épistémologiques

Emmanuelle DANBLON

Ce chapitre à l'ambition de défendre un projet de naturalisation de l'activité rhétorique en cherchant à surmonter une série d'écueils qu'une telle démarche peut occasionner :

- Le réductionnisme (naturaliser, c'est réduire au biologique)
- Le post-modernisme (naturaliser, c'est sortir de sa discipline et utiliser des concepts que l'on ne maîtrise pas ou, pire, que l'on utilise de façon métaphorique)

A ces deux critiques qu'il faut prendre au sérieux, je voudrais répondre de la façon suivante. Tout d'abord, l'attention portée à l'ancrage biologique ne nous condamne pas à nous contenter de ce seul niveau de description, au contraire. Ensuite, le projet de naturalisation de la rhétorique nous engage à une réflexion épistémologique sur le statut de la rhétorique comme discipline. Est-elle une pratique (technique), un art, une science ou une philosophie ? Ce débat conduit à un développement spécifique sur le statut de la raison pratique au regard de la culture de l'artisanat dans laquelle est née la rhétorique avant de proposer une description de la persuasion dans le cadre de cette réflexion générale. Le principal argument est que cette description

de la rhétorique engage à un élargissement du concept de rationalité qui doit tenir compte davantage de l'intelligence pratique. Pour ce faire, les développements actuels en sciences cognitives sont reliés à la description aristotélicienne, elle aussi naturaliste. Enfin, le chapitre conclut pour une pratique sérieuse de l'interdisciplinarité qui offre, à nos yeux, une voie nécessaire au développement humaniste de la recherche.

### A case for emotion awareness

Evgenia PAPAROUNI

The purpose of this paper is to discuss the question whether a boundary can be drawn between a rational and an unreasonable (dysfunctional, or unethical) appeal to emotion.

Data reported by experimental psychologists (as summarized by De Houwer & Hermans 2010; Blanc 2006) show that emotion, far from being a noise in a vacuum of pure thought, interacts with cognition at “low” and “high” levels in the construction of meaning. Such findings contribute to revisiting the notions of bias and subjectivity. Yet literature and our common experience provide a lot of examples where people and peoples have been driven astray by skilful communicators, even if they came to realise it with the benefit of time and hindsight. Beliefs and desires coexist in a set of multiple phenomena such as wishful thinking, self-deception, autosuggestion, stereotypic classifications (Elster 1978, 1983), motivational beliefs (Clément 2006) and, last but not least, in the expression of preferences grounded on values that may lead to self-abnegation and sacrifice; (Livet 2002). The philosophical tradition of dualism and, in some cases, argumentation studies consider it fallacious to rely on emotions. If research stuck to this belief in an ideal speech, it would be at a loss in front of real-life situations. On the opposite side, rhetoric from its very beginnings did study the use of pathos.

Emotions can be reasons to act and can be expressed in a propositional form, which makes them objects of study. They are subjective and situational, a fact that faces us with methodological constraints to be met by case studies from a multidisciplinary perspective. Such an approach paves the way for a Topic of Emotions that may elaborate on the one proposed by Aristotle. Such a Topic may acquire a predictive character, e.g. by linking fear with risk aversion and conservative choice. If we systematically take into account the

specific audience at hand, we can supplement “epistemic vigilance” (Sperber et al. 2010) with “emotion awareness”. In this sense, manipulation could be defined as an appeal to the unconscious, non-“adult” part of personality, through means that fail to keep balance between logos, ethos and pathos (Adam 2002: 46).

### L’argument d’autorité: de sa structure à ses effets

Thierry HERMAN

L’argument d’autorité, schème argumentatif fréquent et occupant une place de choix dans toutes les typologies de types d’argument, est ici revu à l’aune des définitions qu’on lui donne. Partant d’une définition classique articulant une proposition P, un locuteur expert X et la conclusion Q «P est vrai», notre analyse suggère que la conclusion de l’argument d’autorité est moins une question de vérité que d’indubitabilité postulée. Cette force rhétorique de l’argument présenté, qui agit comme si toute mise en doute était exclue, est d’autant plus prégnante que la structure même de l’argument d’autorité ne permet pas de mettre en discussion le propos autorisé. En effet, l’immense majorité des exemples d’argument d’autorité dans les manuels d’argumentation montre que l’autorité est assurée au niveau d’une prémisse au sein d’une argumentation d’un autre type, ce qui sert la tentative d’imposer P à l’allocataire comme allant de soi. Sans considérer le schème comme un sophisme, nous observerons s’il reste un schème identifiable en tant qu’argument d’autorité quand on lui retire les constituants classiques de ce type d’argumentation. Que se passe-t-il si l’expert cité n’est pas présenté comme tel ? Si le locuteur ne prend pas en charge le contenu référentiel de P ? Si P est un fait attesté par un expert et non une opinion ? Si, enfin, le locuteur compte sur sa propre autorité ou présuppose son autorité sans recourir à un tiers expert ou témoin ? Nous interrogeons ces différentes perspectives sur la base d’exemples tirés de la presse écrite, en insistant sur les effets rhétoriques de ce schème plutôt atypique dans son fonctionnement et en montrant l’importance de la modalité épistémique dans l’assertion autoritaire dont le schème de l’argument d’autorité serait une sous-catégorie.

## Argumentation from expert opinion in science journalism: The case of Eureka's Fight Club

Sara GRECO MORASSO

Carlo MORASSO

This paper sets out to analyse fallacious uses of the locus from authority in science journalism based in particular on vague appeals to expert opinion. To do so, we adopt a case-study approach (Flyvbjerg 2001), focusing our analysis on the column *Fight Club* published by the monthly scientific supplement *Eureka* to the British newspaper *The Times* (issues published from October 2009 to November 2010). We analyse scientific journalism as a context of argumentation (Rigotti and Rocci 2006, van Eemeren 2010), considering its goals and the roles fulfilled by scientists, journalists and the public (Lesnher 2003, Bubela et al. 2009). A sound approach to expert opinion is crucial in this context, as laypersons have normally no choice but trusting scientific journalists to understand data which would otherwise be inaccessible to them but which can bear important consequences for their present and future lives.

From the theoretical point of view, we start from Douglas Walton's well-documented approach to appeals to expert opinion (Walton 1997). Yet we extend and complement this approach in two important respects. First, we follow Goodwin's (2010) suggestion to reinterpret the expert-layperson relationship as an agency relationship. The layperson is the principal in a relationship in which she needs to retain the agent, i.e. the expert, to do something she cannot do for herself. Goodwin's hypothesis inserts appeals to expert opinion in the bigger picture of a human relationship in which both principal and agent have their own rhetorical goals, which sometimes may not be aligned, thus explaining possible origins of fallacious uses of the locus from expert opinion as derailments of the arguers' strategic manoeuvring (van Eemeren and Houtlosser 2002).

Second, we associate Walton's critical questions to the premises of the argument scheme from expert opinion on the basis of the Argumentum Model of Topics (AMT, Rigotti and Greco Morasso 2011). This allows understanding the kind of premises which vague uses of expert opinion are associated to. What results from our analysis is the weakness of the material component of this argument scheme. From the cognitive point of view, the interest in the locus from expert opinion lays in the fact that, when it is used in a vague way, it may become a sort of cognitive immunisation against



argumentation; in fact, appealing to the authority of experts in argumentative discussions may be used to invite to suspend one's critical attitude towards his or her decision.

In sum, our paper presents advances on the argument scheme from expert opinion at two levels. At the theoretical level, it contributes to discussion on this specific argument scheme and the cognitive implications of its possible fallacious uses. At the level of the study of argumentation in context, it sheds some light on the debate on science journalism, which is nowadays fiercely open among scientists and the wider public.

### Two-sided rhetorical strategies in top management's letters to shareholders and stakeholders from corporate reports

Ioana Agatha FILIMON

This study explores the use of two-sided rhetorical strategies in two corpora of introductory letters to shareholders, respectively, stakeholders from annual and corporate social responsibility reports. By means of a mixed methodological approach that combines the analytical and evaluative tools of argumentation theory (van Eemeren & Grootendorst 2004, Rigotti & Greco Morasso 2010) with insights coming from the experimental research on message sidedness effects (O'Keefe 1999), the study observes the extent of use, the structure, and the potential impact of the two-sided rhetorical strategies on the communicative effectiveness of the two types of introductory letters (i.e., on the capacity of these letters to provide a credible and persuasive account on the corporate activity). Although the study does not empirically evaluate the persuasiveness and the credibility of the two-sided rhetorical strategies identified in the corpus letters, it does try to make a coarse estimation of their impact, *ex hypothesi*, by re-interpreting the results of the argumentative analysis in the light of O'Keefe's (1999) meta-analytical findings and hypotheses about the message sidedness effects in various communicative contexts.

### Presupposing redefinitions

Fabrizio MACAGNO

Redefinitions are crucial communicative moves in argumentation. The commonly accepted meaning of a word is replaced by a new one, which is

used to refer to a new concept or fragment of reality. The value judgment commonly associated with the old meaning often survives the redefinition, and can be used by the speaker to influence the interlocutor's assessment of the new referent. These moves can be extremely dangerous when left implicit and used to introduce ambiguity and take advantage of the equivocation between the accepted definition and the new one. On this view, the concept used to classify reality is different from the one applied for evaluating it. Stevenson analyzed the argumentative use of redefinitions describing persuasive definitions. They are definitions aimed at altering the descriptive content of a word (its denotation) leaving unaltered its emotive meaning. However, Stevenson did not provide any criteria to assess when and why these moves are fallacious. The purpose of this research is to inquire into redefinitions by distinguishing between the propositional, the pragmatic, and the argumentative level. Redefinitions are analyzed as dialogical moves or speech acts used to reach a value judgment. When left implicit, redefinitions may be unduly presumed to be accepted or acceptable. As a consequence, the interlocutor is led to accept unwanted conclusions, or to fulfill the burden of proving that such definition is not commonly accepted.

### Présuppositions discursives, assertion d'arrière-plan et persuasion

Louis DE SAUSSURE

La littérature cognitive a documenté une série de biais cognitifs (ou illusions cognitives), i.e. des heuristiques de traitement de l'information spécifiquement génératrices d'erreurs (cf. Pohl 2004 pour un panorama). Nous suggérons que ces biais peuvent s'interpréter, en les croisant dans un autre cadre disciplinaire – la linguistique –, comme directement produits par des effets de présupposition. Plus précisément, nous proposons de montrer en quoi ces biais provoquent des erreurs dans le processus d'attribution de pertinence (au sens de Sperber & Wilson 1995) aux informations présupposées. Nous suggérerons que la présupposition, par nature, échappe à l'extension du contexte de recherche de pertinence, articulant ainsi notre analyse avec le modèle de Maillat & Oswald (2009), pour qui la persuasion repose essentiellement sur des principes de restriction d'accès au contexte.

Les présuppositions résistent à l'enchaînement discursif (Ducrot 1972) et sont réputées déjà admises dans l'arrière-plan conversationnel (Stalnaker 1978, Merin 2003). De ce fait, les présuppositions résistent normalement à

une évaluation critique approfondie, sauf hypothèses contraires fortes. Les présuppositions, suggérons-nous, sont interprétées selon des routines de traitement superficiel (*shallow processing*): elles se trouvent automatiquement intégrées comme consistantes et valides. Cependant, il est également connu que les présuppositions sont seulement présentées comme déjà connues alors qu'elles peuvent ne pas l'être en l'espèce, et peuvent à leur tour entrer dans des inférences et intervenir dans des processus heuristiques. Admettre les présuppositions répond à un impératif cognitif simple de pertinence qui consiste à économiser l'évaluation du «déjà connu». De ce fait, elles comportent un risque critique si l'information qu'elles communiquent sont impliquées dans la satisfaction d'objectifs persuasifs : elles ont plus de chances que l'information posée de contourner l'évaluation critique.

Nous suggérons qu'un nombre important de biais documentés dans la littérature psychologique ont directement à voir avec la présupposition : cadrage (Tversky & Kahneman 1974), illusions cognitives de type illusion de Moïse (Reder & Kusbit 1991); les effets de label (Loftus & Palmer 1974), et de mésinformation (Loftus, Miller & Burns 1978) déclenchent des présuppositions encyclopédiques ou existentielles (typiquement avec les descriptions définies). Les biais de confirmation (i.e. la tendance à confirmer les hypothèses préexistantes) peut également s'interpréter comme une forme de biais présuppositionnel relatif au «déjà su», ce qui conduit à suggérer que l'observation de Stalnaker devrait s'étendre au-delà des connaissances «partagées» (common ground).

Après avoir détaillé ce qui précède, nous analyserons le cas spécifique du récent vote de la Suisse contre la construction des minarets par l'effet de pertinence associée à des présuppositions pragmatiques déclenchées non seulement par l'énoncé et son contexte mais spécifiquement par leur valeur présumée pertinente : pour que la question posée aux votants («acceptez-vous d'interdire la construction de minarets») soit pertinente, il faut qu'un certain nombre de conditions soient réunies, qui ont été prises pour acquises par les destinataires alors qu'elles n'avaient pas lieu d'être.

## Pragmatics, cognitive heuristics and the straw man fallacy

Steve OSWALD

Marcin LEWIŃSKI

This paper presents a contextually and cognitively grounded account of the straw man fallacy by addressing two interrelated questions: (i) when can we justifiably say that the straw man occurred? and (ii) how can we explain its remaining covert, its persuasiveness and, in some cases, its manipulativeness? To answer both these questions, we draw on the integrated pragma-dialectical theory of argumentation focused on the concept of strategic manoeuvring (van Eemeren & Houtlosser 1999) and on cognitive pragmatics (cf. Sperber & Wilson 1995) and show that they can consistently be integrated in a comprehensive account of fallaciousness that combines dialectical, rhetorical and pragmatic insights.

The main contribution of this chapter focuses on this second question and lies in its attempt at building bridges between cognitive science and rhetoric. To this end, it considers the role information processing mechanisms may play in argumentative contexts (and fallacious ones in particular). Chiefly, the research presented here considers the role of cognitive biases and heuristics in the success of argumentatively fallacious discourse, that is, in rhetorically significant contexts. The authors defend the assumption that fallacies may be rhetorically effective when they manage to circumvent epistemic vigilance filters, either by managing to divert attention from critical information or by making critical information perceived as irrelevant by the addressee.

More specifically, from a cognitive pragmatic perspective, the authors propose, following Oswald (2010, 2011, *forth.*), that fallacies can be viewed as an arguer's attempt at imposing contextual constraints on addressees' interpretations, the success of which should characteristically exclude information about the fallacious nature of the argument. They accordingly formulate their answer to question (ii) in terms of information processing mechanisms governing utterance interpretation and show, through extensive analysis of a concrete example taken from political discourse, that the straw man operates a constraint that leaves critical information out of the meaning derivation procedure.

Favoriser la confrontation interdisciplinaire et internationale de toutes les formes de recherches consacrées à la communication humaine, en publiant sans délai des travaux scientifiques d'actualité: tel est le rôle de la collection *Sciences pour la communication*. Elle se propose de réunir des études portant sur tous les langages, naturels ou artificiels, et relevant de toutes les disciplines sémiologiques: linguistique, psychologie ou sociologie du langage, sémiotiques diverses, logique, traitement automatique, systèmes formels, etc. Ces textes s'adressent à tous ceux qui voudront, à quelque titre que ce soit et où que ce soit, se tenir au courant des développements les plus récents des sciences du langage.

Ouvrages parus

SC!

1. Alain Berrendonner – L'éternel grammairien Etude du discours normatif, 1982 (épuisé)
2. Jacques Moeschler – Dire et contredire Pragmatique de la négation et acte de réfutation dans la conversation, 1982 (épuisé)
3. C. Bertaux / J.-P. Desclés / D. Dubarle / Y. Gentilhomme / J.-B. Grize / I. Mel'cuk / P. Scheurer / R. Thom – Linguistique et mathématiques · Peut-on construire un discours cohérent en linguistique? · Table ronde organisée par l'ATALA, le Séminaire de philosophie et mathématiques de l'Ecole Normale Supérieure de Paris et le Centre de recherches sémiologiques de Neuchâtel (Neuchâtel, 29-31 mai 1980), 1982
4. Marie-Jeanne Borel / Jean-Blaise Grize / Denis Miéville – Essai de logique naturelle, 1983, 1992
5. P. Bange / A. Bannour / A. Berrendonner / O. Ducrot / J. Kohler-Chesny / G. Lüdi / Ch. Perelman / B. Py / E. Roulet – Logique, argumentation, conversation · Actes du Colloque de pragmatique (Fribourg, 1981), 1983
6. Alphonse Costadau: Traité des signes (tome I) – Edition établie, présentée et annotée par Odile Le Guern-Forel, 1983
7. Abdelmajid Ali Bouacha – Le discours universitaire · La rhétorique et ses pouvoirs, 1984
8. Maurice de Montmollin – L'intelligence de la tâche · Eléments d'ergonomie cognitive, 1984, 1986 (épuisé)
9. Jean-Blaise Grize (éd.) – Sémiologie du raisonnement · Textes de D. Apothélos, M.-J. Borel, J.-B. Grize, D. Miéville, C. Péquegnat, 1984
10. Catherine Fuchs (éd.) – Aspects de l'ambiguïté et de la paraphrase dans les langues naturelles Textes de G. Bès, G. Boulakia, N. Catach, F. François, J.-B. Grize, R. Martin, D. Slakta, 1985
11. E. Roulet / A. Auchlin / J. Moeschler / C. Rubattel / M. Schelling – L'articulation du discours en français contemporain, 1985, 1987, 1991 (épuisé)
12. Norbert Dupont – Linguistique du détachement en français, 1985
13. Yves Gentilhomme – Essai d'approche microsystemique · Théorie et pratique · Application dans le domaine des sciences du langage, 1985
14. Thomas Bearth – L'articulation du temps et de l'aspect dans le discours toura, 1986
15. Herman Parret – Prolégomènes à la théorie de l'énonciation · De Husserl à la pragmatique, 1987
16. Marc Bonhomme – Linguistique de la métonymie · Préface de M. Le Guern, 1987 (épuisé)
17. Jacques Rouault – Linguistique automatique · Applications documentaires, 1987
18. Pierre Bange (éd.) – L'analyse des interactions verbales: «La dame de Caluire. Une consultation» · Actes du Colloque tenu à l'Université Lyon II (13-15 décembre 1985), 1987
19. Georges Kleiber – Du côté de la référence verbale · Les phrases habituelles, 1987

20. Marianne Kilani-Schoch – Introduction à la morphologie naturelle, 1988
21. Claudine Jacquenod – Contribution à une étude du concept de fiction, 1988
22. Jean-Claude Beacco – La rhétorique de l'historien · Une analyse linguistique du discours, 1988
23. Bruno de Foucault – Les structures linguistiques de la genèse des jeux de mots, 1988
24. Inge Egner – Analyse conversationnelle de l'échange réparateur en wobé · Parler WEE de Côte d'Ivoire, 1988
25. Daniel Peraya – La communication scalène · Une analyse sociosémiotique de situations pédagogiques, 1989
26. Christian Rubattel (éd.) – Modèles du discours · Recherches actuelles en Suisse romande · Actes des Rencontres de linguistique française (Crêt-Bérard, 1988), 1989
27. Emilio Gattico – Logica e psicologia · Studi piagetiani e postpiagetiani, 1989
28. Marie-José Reichler-Béguelin (éd.) – Perspectives méthodologiques et épistémologiques dans les sciences du langage · Actes du Colloque de Fribourg (11-12 mars 1988), 1989
29. Pierre Dupont – Eléments logico-sémantiques pour l'analyse de la proposition, 1990
30. Jacques Wittwer – L'analyse relationnelle · Une physique de la phrase écrite · Introduction à la psychosyntaxique, 1990
31. Michel Chambréuil / Jean-Claude Pariente – Langue naturelle et logique · La sémantique intentionnelle de Richard Montague, 1990
32. Alain Berrendonner / Herman Parret (éds) – L'interaction communicative, 1990 (épuisé)
33. Jacqueline Bideaud / Olivier Houdé – Cognition et développement · Boîte à outils théoriques · Préface de Jean-Blaise Grize, 1991 (épuisé)
34. Beat Münch – Les constructions référentielles dans les actualités télévisées · Essai de typologie discursive, 1992
35. Jacques Theureau – Le cours d'action · Analyse sémio-logique · Essai d'une anthropologie cognitive située, 1992 (épuisé)
36. Léonardo Pinsky (†) – Concevoir pour l'action et la communication · Essais d'ergonomie cognitive · Textes rassemblés par Jacques Theureau et collab., 1992
37. Jean-Paul Bernié – Raisonner pour résumer · Une approche systémique du texte, 1993
38. Antoine Auchlin – Faire, montrer, dire – Pragmatique comparée de l'énonciation en français et en chinois, 1993
39. Zlatka Guentcheva – Thématization de l'objet en bulgare, 1993
40. Corinne Rossari – Les opérations de reformulation · Analyse du processus et des marques dans une perspective contrastive français – italien, 1993, 1997
41. Sophie Moirand / Abdelmadjid Ali Bouacha / Jean-Claude Beacco / André Collinot (éds) – Parcours linguistiques de discours spécialisés · Colloque en Sorbonne les 23-24-25 septembre 1992, 1994, 1995

42. Josiane Boutet – Construire le sens · Préface de Jean-Blaise Grize, 1994, 1997
43. Michel Goyens – Emergence et évolution du syntagme nominal en français, 1994
44. Daniel Duprey – L'universalité de «bien» · Linguistique et philosophie du langage, 1995
45. Chantal Rittaud-Hutinet – La phonopragmatique, 1995
46. Stéphane Robert (éd.) – Langage et sciences humaines: propos croisés · Actes du colloque «Langues et langages» en hommage à Antoine Culioli (Ecole normale supérieure, Paris, 11 décembre 1992), 1995
47. Gisèle Holtzer – La page et le petit écran: culture et télévision · Le cas d'Apostrophes, 1996
48. Jean Wirtz – Métadiscours et déceptivité · Julien Torma vu par le Collège de 'Pataphysique, 1996
49. Vlad Alexandrescu – Le paradoxe chez Blaise Pascal · Préface de Oswald Ducrot, 1997
50. Michèle Grossen, Bernard Py (éds) – Pratiques sociales et médiations symboliques, 1997
51. Daniel Luzzati / Jean-Claude Beacco / Reza Mir-Samii / Michel Murat / Martial Vivet (éds) – Le Dialogique · Colloque international sur les formes philosophiques, linguistiques, littéraires, et cognitives du dialogue (Université du Maine, 15-16 septembre 1994), 1997
52. Denis Miéville / Alain Berrendonner (éds) – Logique, discours et pensée · Mélanges offerts à Jean-Blaise Grize, 1997, 1999
53. Claude Guimier (éd.) – La thématization dans les langues · Actes du colloque de Caen, 9-11 octobre 1997, 1999, 2000
54. Jean-Philippe Babin – Lexique mental et morphologie lexicale, 1998, 2000
55. Thérèse Jeanneret – La coénonciation en français · Approches discursive, conversationnelle et syntaxique, 1999
56. Pierre Boudon – Le réseau du sens · Une approche monadologique pour la compréhension du discours, 1999 (épuisé)
58. Jacques Moeschler, Marie-José Béguélin (éds) – Référence temporelle et nominale. Actes du 3<sup>e</sup> cycle romand de Sciences du langage, Cluny (15–20 avril 1996), 2000
59. Henriette Gezundhajt – Adverbes en *-ment* et opérations énonciatives · Analyse linguistique et discursive, 2000
60. Christa Thomsen – Stratégies d'argumentation et de politesse dans les conversations d'affaires · La séquence de requête, 2000
61. Anne-Claude Berthoud, Lorenza Mondada (éds) – Modèles du discours en confrontation, 2000
62. Eddy Roulet, Anne Grobet, Laurent Filliettaz, avec la collaboration de Marcel Burger – Un modèle et un instrument d'analyse de l'organisation du discours, 2001
63. Annie Kuyumcuyan – Diction et mention · Pour une pragmatique du discours narratif, 2002



64. Patrizia Giuliano – La négation linguistique dans l’acquisition d’une langue étrangère · Un débat conclu? 2004
65. Pierre Boudon – Le réseau du sens II · Extension d’un principe monadologique à l’ensemble du discours, 2002
66. Pascal Singy (éd.) – Le français parlé dans le domaine francoprovençal · Une réalité plurinationale, 2002
67. Violaine de Nuchèze, Jean-Marc Colletta (éds) – Guide terminologique pour l’analyse des discours · Lexique des approches pragmatiques du langage, 2002
68. Hanne Leth Andersen, Henning Nølke – Macro-syntaxe et macro-sémantique · Actes du colloque international d’Århus, 17-19 mai 2001, 2002
69. Jean Charconnet – Analogie et logique naturelle · Une étude des traces linguistiques du raisonnement analogique à travers différents discours, 2003
70. Christopher Laenzlinger – Initiation à la Syntaxe formelle du français · Le modèle *Principes et Paramètres* de la Grammaire Générative Transformationnelle, 2003
71. Hanne Leth Andersen, Christa Thomsen (éds) – Sept approches à un corpus · Analyses du français parlé, 2004
72. Patricia Schulz – Description critique du concept traditionnel de «métaphore», 2004
73. Joël Gapany – Formes et fonctions des relatives en français · Etude syntaxique et sémantique, 2004
74. Anne Catherine Simon – La structuration prosodique du discours en français · Une approche multidimensionnelle et expérientielle, 2004
75. Corinne Rossari, Anne Beaulieu-Masson, Corina Cojocariu, Anna Razgouliaeva – Autour des connecteurs · Réflexions sur l’énonciation et la portée, 2004
76. Pascal Singy (éd.) – Identités de genre, identités de classe et insécurité linguistique, 2004
77. Liana Pop – La grammaire graduelle, à une virgule près, 2005
78. Injoo Choi-Jonin, Myriam Bras, Anne Dagnac, Magali Rouquier (éds) – Questions de classification en linguistique: méthodes et descriptions · Mélanges offerts au Professeur Christian Molinier, 2005
79. Marc Bonhomme – Le discours métonymique, 2005
80. Jasmina Milićević – La paraphrase · Modélisation de la paraphrase langagière, 2007
81. Gilles Siouffi, Agnès Steuckardt (éds) – Les linguistes et la norme · Aspects normatifs du discours linguistique, 2007
82. Agnès Celle, Stéphane Gresset, Ruth Huart (éds) – Les connecteurs, jalons du discours, 2007
83. Nicolas Pepin – Identités fragmentées · Eléments pour une grammaire de l’identité, 2007
84. Olivier Bertrand, Sophie Prévost, Michel Charolles, Jacques François, Catherine Schnedecker (éds) – Discours, diachronie, stylistique du français · Etudes en hommage à Bernard Combettes, 2008

85. Sylvie Mellet (dir.) – Concession et dialogisme · Les connecteurs concessifs à l'épreuve des corpus, 2008
86. Benjamin Fagard, Sophie Prévost, Bernard Combettes, Olivier Bertrand (éds) – Evolutions en français · Etudes de linguistique diachronique, 2008
87. Denis Apothéloz, Bernard Combettes, Franck Neveu (éds) – Les linguistiques du détachement · Actes du colloque international de Nancy (7-9 juin 2006), 2009
88. Aris Xanthos – Apprentissage automatique de la morphologie · Le cas des structures racine–schème, 2008
89. Bernard Combettes, Céline Guillot, Evelyne Oppermann-Marsaux, Sophie Prévost, Amalia Rodríguez Somolinos (éds) – Le changement en français · Etudes de linguistique diachronique, 2010
90. Camino Álvarez Castro, Flor M<sup>a</sup> Bango de la Campa, María Luisa Donaire (éds.) – Liens linguistiques · Etudes sur la combinatoire et la hiérarchie des composants, 2010
91. Marie-José Béguelin, Mathieu Avanzi, Gilles Corminboeuf (éds) – La Parataxe · Entre dépendance et intégration; Tome 1, 2010
92. Marie-José Béguelin, Mathieu Avanzi, Gilles Corminboeuf (éds) – La Parataxe · Structures, marquages et exploitations discursives; Tome 2, 2010
93. Nelly Flaux, Dejan Stosic, Co Vet (éds) – Interpréter les temps verbaux, 2010
94. Christian Plantin – Les bonnes raisons des émotions · Principes et méthode pour l'étude du discours *émotionné*, 2011
95. Dany Amiot, Walter De Mulder, Estelle Moline et Dejan Stosic (éds) – *Ars Grammatica* · Hommages à Nelly Flaux, 2011.
96. André Horak (éd.) – La litote · Hommage à Marc Bonhomme, 2011.
97. Franck Neveu, Nicole Le Querler et Peter Blumenthal (éds) – Au commencement était le verbe. Syntaxe, sémantique et cognition · Mélanges en l'honneur du Professeur Jacques François, 2011.
98. Louis de Saussure et Alain Rihs (éds) – Etudes de sémantique et pragmatique françaises, 2012.
99. L. de Saussure, A. Borillo, M. Vuillaume (éds) – Grammaire, lexique, référence. Regards sur le sens · Mélanges offerts à Georges Kleiber pour ses quarante ans de carrière, 2012.
100. Groupe de Fribourg – Grammaire de la période, 2012
101. C. Guillot, B. Combettes, A. Lavrentiev, E. Oppermann-Marsaux, S. Prévost (éd.) – Le changement en français · Etudes de linguistique diachronique, 2012.
102. Gudrun Vanderbauwhede – Le déterminant démonstratif en français et en néerlandais · Théorie, description, acquisition, 2012.
103. Genoveva Puskás – Initiation au Programme Minimaliste · Eléments de syntaxe comparative, 2013.
104. Coco Norén, Kerstin Jonasson, Henning Nølke et Maria Svensson (éds) – Modalité, évidentialité et autres friandises langagières · Mélanges offerts à Hans Kronning à l'occasion de ses soixante ans, 2013.

105. Jean-Claude Anscombe, María Luisa Donaire, Pierre Patrick Haillet (éds.) – Opérateurs discursifs du français · Eléments de description sémantique et pragmatique, 2013.
106. Laurent Gosselin, Yann Mathet, Patrice Enjalbert, Gérard Becher – Aspects de l'itération · L'expression de la répétition en français: analyse linguistique et formalisation, 2013.
107. Alain Rihs – Subjonctif, gérondif et participe présent en français · Une pragmatique de la dépendance verbale, 2013.
108. Emmanuelle Labeau & Jacques Bres (éds.) – Evolution in Romance Verbal Systems, 2013.
109. A paraître
110. Christiane Soum-Favaro, Annelise Coquillon & Jean-Pierre Chevrot (éds.) – La liaison: approches contemporaines, 2014.
111. A paraître
112. Thierry Herman et Steve Oswald (éds) – Rhétorique et cognition / Rhetoric and Cognition, 2014.
113. Giovanni Gobber Andrea Rocci – Language, reason and education, 2014 · Studies in honor of Eddo Rigotti, 2014.