

## PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/74866>

Please be advised that this information was generated on 2017-12-06 and may be subject to change.

## **The good, the bad and the persuasive**

Normative quality and actual persuasiveness of  
arguments from authority, arguments from cause to effect and  
arguments from example

Published by  
LOT  
Janskerkhof 13  
3512 BL Utrecht  
The Netherlands

phone: +31 30 253 6006  
fax: +31 30 253 6406  
e-mail: [lot@let.uu.nl](mailto:lot@let.uu.nl)  
<http://www.lotschool.nl>

ISBN: 978-94-6093-015-7  
NUR 616

Copyright © 2010: Ester Šorm. All rights reserved.

# **The good, the bad and the persuasive**

Normative quality and actual persuasiveness of  
arguments from authority, arguments from cause to effect and  
arguments from example

Een wetenschappelijke proeve op het gebied van de Letteren

## **Proefschrift**

ter verkrijging van de graad van doctor  
aan de Radboud Universiteit Nijmegen  
op gezag van de rector magnificus prof. mr. S.C.J.J. Kortmann,  
volgens besluit van het College van Decanen  
in het openbaar te verdedigen op dinsdag 23 februari 2010  
om 13.30 uur precies

door

**Ester Šorm**  
geboren op 3 februari 1981  
te Hilversum

Promotores:

prof. dr. J.A.L. Hoeken  
prof. dr. P.J.M.C. Schellens

Manuscriptcommissie:

prof. dr. J.W.J. Beentjes  
prof. dr. F.H. van Eemeren (Universiteit van Amsterdam)  
prof. dr. M.J.P. van Mulken  
prof. dr. W.P.M.S. Spooren (Vrije Universiteit Amsterdam)  
prof. dr. D.H.J. Wigboldus

## Acknowledgements

The research carried out for this dissertation was sponsored by the Netherlands Organisation for Scientific Research (NWO). I am grateful to the Organisation for the funding provided.

Hans Hoeken and Peter Jan Schellens supervised the project. I am enormously grateful for their comments, advice and commitment to the project, and to me as PhD. I particularly appreciate their continuing helpfulness and encouragement, and also the excellent feedback I received on my drafts. It was a pleasure to be guided by such outstanding promotores, who I consider to meet all criteria for good cognitive authorities and good normative examples to all other promotores. I would like to extend my thanks to Lettica Hustinx for advising the research team.

Rian Timmers was my project partner, with whom I collaborated over the years. I am deeply grateful for her contributions to the research and numerous significant suggestions. We have not only had constructive conversations about scientific topics, but also cheerful chats about more private matters. I was extremely fortunate to have a companion like her.

I would also like to express my gratitude to all my wonderful colleagues at the department of Business Communication Studies and the department of Dutch Language and Culture. It was the latter department where I had my working desk, my computer and the nicest roommate one can imagine: Sjaak Kroon. Although we only met on Fridays, I will never forget the inspiring conversations and his sincere interest in my work and private life.

There is another colleague who should be mentioned here: Christian Burgers. I thank him for his outstanding collegiality and support. He was always there for me.

I owe a debt to my mother, who listened to my extended and uninterrupted speeches on joy and sorrow at work. With her amazing ability to put things into perspective, things stayed in the contexts where they belonged.

Many other people contributed to the research. Many thanks to Liesbeth van der Brugge, and other staff members of the Centrale Bibliotheek De Mariënborg in Nijmegen. They provided a pleasant research setting, one to which I could invite participants for interviews and filling out questionnaires. I am also much obliged to Floris Rooijmans and Britt van Enthoven for their professional and enthusiastic assistance during the experimental research. I would also like to thank Lidwien Cluitmans for being a second coder.

Thanks to her work, reliability could be tested. I would also like to thank Yvette Linders, Joske Duifhuizen, Jet Meerkerk, Lindsey Hubers and Christine Liebrecht for their assistance in designing and carrying out the preliminary experimental study. A big thank you to all students of Business Communication at the Radboud University Nijmegen who were willing to conduct focus groups and interviews, and to recruit participants. I thank Jos Hornikx who allowed me to use claims in the experimental material that he had used in his own research. I would like to thank two anonymous reviewers of the Tijdschrift voor Taalbeheersing for their valuable comments on a draft of an article that I adapted into a chapter for this dissertation. I am extremely thankful to the UTN corrector, who corrected the non-native English in this dissertation. Thanks to all who acted as participants in the empirical research.

I am particularly grateful to someone who evolved from a sympathetic colleague to a dear and loving husband in less time than it took to write this dissertation: Gijsbert Rutten. I benefited from his scientific knowledge and skills, his sense of humour and unconditional support. I feel privileged to have met him.

Brussels, June 2009

## Contents

<b>1</b>	<b>Introduction</b>	<b>11</b>
1.1	Argument quality in persuasion research	11
1.2	Argument quality in argumentation studies	15
1.3	Pragmatic arguments: an introduction	20
1.4	Dimensions of a pragmatic argument: probability and desirability of consequences	22
1.5	The quality of pragmatic arguments	24
1.6	Types of argument supporting the probability statement	26
1.7	Aim of research and research questions	28
1.8	Overview of the study	29
<b>PART I: CRITERIA FOR ARGUMENT QUALITY</b>		<b>31</b>
<b>2</b>	<b>The quality of the argument from authority</b>	<b>33</b>
2.1	Introduction	33
2.1.1	Types of authority	33
2.1.2	Epistemic authority	35
2.2	Argumentation-theoretical criteria	39
2.2.1	Method	40
2.2.2	Results	41
2.2.3	Conclusion and discussion	48
2.3	Laymen criteria	50
2.3.1	Methodological study	50
2.3.2	What criteria do laymen use?	53
2.3.2.1	Method	54
2.3.2.2	Results	62
2.4	Conclusion and discussion	72
<b>3</b>	<b>The quality of the argument from cause to effect</b>	<b>77</b>
3.1	Introduction	77
3.1.1	Causality conditions	77
3.1.2	The argument from cause to effect	79
3.2	Argumentation-theoretical criteria	81
3.2.1	Method	81
3.2.2	Results	82
3.2.3	Conclusion and discussion	91
3.3	Laymen criteria	93
3.3.1	Method	93
3.3.2	Results	102
3.4	Conclusion and discussion	118



<b>4</b>	<b>The quality of the argument from example</b>	<b>123</b>
4.1	Introduction	123
	4.1.1 Functions of an example	123
	4.1.2 Sampling principles in scientific methodology	128
4.2	Argumentation-theoretical criteria	130
	4.2.1 Method	132
	4.2.2 Results	132
	4.2.3 Conclusion and discussion	140
4.3	Laymen criteria	141
	4.3.1 Method	141
	4.3.2 Results	146
4.4	Conclusion and discussion	166
<b>PART II: THE RELATION BETWEEN ARGUMENT QUALITY AND ACTUAL PERSUASIVENESS: AN EXPERIMENTAL APPROACH</b>		<b>173</b>
<b>5</b>	<b>The perceived reasonableness and actual persuasiveness of normatively strong and normatively weak arguments</b>	<b>175</b>
5.1	Introduction	175
5.2	Method	179
	5.2.1 Material	179
	5.2.2 Instrumentation	184
	5.2.3 Design	185
	5.2.4 Respondents	185
	5.2.5 Procedure	185
5.3	Results	186
	5.3.1 Argument from authority	186
	5.3.2 Argument from cause to effect	188
5.4	Conclusion and discussion	190
<b>6</b>	<b>The actual persuasiveness of normatively strong and weak arguments from authority, arguments from cause to effect and arguments from example</b>	<b>193</b>
6.1	Introduction	193
6.2	Method	199
	6.2.1 Pre-test	199
	6.2.2 Material	200
	6.2.3 Respondents	206
	6.2.4 Design	207

6.2.5	Instrumentation	207
6.2.6	Procedure	211
6.2.7	Statistical tests	211
6.3	Results	212
6.3.1	Manipulation checks	212
6.3.2	Target items	215
6.4	Conclusion and discussion	221
<b>7</b>	<b>General conclusion and discussion</b>	<b>227</b>
7.1	Conclusion	227
7.2	Discussion	231
7.2.1	Explanations	231
7.2.2	Implications for argumentation studies and persuasion research	232
7.3	Limitations and suggestions for further research	234
	<b>References</b>	<b>241</b>
	<b>Appendices chapter 2</b>	
2A	Material used for uncovering laymen criteria for the argument from authority: open case with desirability claim	251
2B	Material used for uncovering laymen criteria for the argument from authority: open case with probability claim	252
2C	Material used for uncovering laymen criteria for the argument from authority: closed case with desirability claim	253
2D	Material used for uncovering laymen criteria for the argument from authority: closed case with probability claim	254

### **Appendices chapter 3**

3A	Material used for uncovering laymen criteria for the argument from cause to effect: open case with claim 1	255
3B	Material used for uncovering laymen criteria for the argument from cause to effect: open case with claim 2	256
3C	Material used for uncovering laymen criteria for the argument from cause to effect: closed case with claim 3	257
3D	Material used for uncovering laymen criteria for the argument from cause to effect: closed case with claim 4	258

### **Appendices chapter 4**

4A	Material used for uncovering laymen criteria for the argument from example: open case with claim 1	259
4B	Material used for uncovering laymen criteria for the argument from example: open case with claim 2	260
4C	Material used for uncovering laymen criteria for the argument from example: closed case with claim 3	261
4D	Material used for uncovering laymen criteria for the argument from example: closed case with claim 4	262

### **Appendices chapter 6**

6A	Pre-test material and results	263
6B	Experimental material	269
6C	Results manipulation checks	280

<b>Summary in Dutch</b>	<b>284</b>
-------------------------	------------

<b>Curriculum Vitae</b>	<b>292</b>
-------------------------	------------

# 1. Introduction

## 1.1 Argument quality in persuasion research

Persuasion can be described as “a successful intentional effort at influencing another’s mental state through communication in a circumstance in which the persuadee has some measure of freedom” (O’Keefe, 2002, p. 5). It is used to implement attitude change. Campaigns have been launched to change the way people think about a variety of issues, varying from substance abuse to fire risk. Politicians attempt to influence opinions toward their policies. In advertising, companies are trying to sell their products by changing the ways in which consumers evaluate their products. The study of persuasion and attitude change is not new; it has fascinated rhetoricians for centuries and up till now, scholars have been interested in the question which message components determine the impact of a persuasive attempt<sup>1</sup>. Many empirical and theoretical studies have been conducted to find out how various message characteristics influence persuasion processes (see, for reviews, O’Keefe, 2002; Perloff, 2003).

Persuasion research is concerned with discovering the factors that influence the effectiveness of persuasive messages and providing explanations for these effects. Probably the most influential approach to persuasion is offered by the Elaboration Likelihood Model (ELM) developed by Petty and Cacioppo (1986a, 1986b)<sup>2</sup>. The ELM<sup>3</sup> suggests that under different circumstances, people vary in the degree to which they are likely to engage in careful thinking (termed ‘elaboration’) about message arguments. Sometimes they will consider the message arguments attentively and sometimes they will take refuge in less effortful thinking, relying on heuristics such as ‘credible sources are usually right’ or ‘if many people like a movie, it is probably good’. Depending on the degree of elaboration, two different kinds of persuasion processes can be involved. The *central route* represents the persuasion process involved when elaboration is

---

<sup>1</sup> For a review of historical and contemporary persuasion scholarship, see Perloff (2003, pp. 3-35).

<sup>2</sup> There is a variety of alternative perspectives on persuasion. For a recent overview of these perspectives, I refer to O’Keefe (in press). I will focus on the ELM in this dissertation, as this is probably the most influential theory of persuasion.

<sup>3</sup> The following characterization of the ELM is partially borrowed from O’Keefe (2008). See O’Keefe (2002, pp. 137-168) for a more detailed treatment of the Elaboration Likelihood Model.

relatively high and the *peripheral route* represents the persuasion process involved when elaboration is relatively low. As the ELM proposes that persuasion can occur as a result of two different processes, it is called a dual process-theory.

A number of factors can influence the degree of elaboration and hence influence which route to persuasion is activated. These factors can influence either elaboration motivation or elaboration ability. Elaboration motivation can be influenced by factors such as the degree of topic involvement (the extent to which the topic is personally relevant to the receiver) or the tendency to engage in careful thinking, generally referred to by 'need for cognition'. Elaboration ability can be influenced by factors such as the degree of distraction or relevant background knowledge. For example, when someone is involved in the topic, enjoys careful thinking, is not distracted or has prior knowledge of the subject matter, elaboration is likely to be high and the message will be processed along the central route. However, when someone is uninvolved, does not like careful thinking, is distracted or lacks relevant prior knowledge, elaboration is likely to be low and the message will be processed along the peripheral route.

Although many studies have approached the Elaboration Likelihood Model theoretically and empirically, the concept of central processing still lacks a clear definition. Petty and Cacioppo remain somewhat vague in their descriptions of central processing, as illustrated by the following citation:

"One [route to attitude change], called the *central route*, says that attitude change results from a person's careful consideration of information that reflects what that person feels are the true merits of a particular attitudinal position. According to this view, if under scrutiny the message arguments are found to be cogent and compelling, favorable thoughts will be elicited that will result in attitude change in the direction of the advocacy. If the arguments are found to be weak and specious, they will be counterargued and the message will be resisted – or boomerang (change opposite to that intended) may even occur." (1984, p. 70).

It becomes clear from Petty and Cacioppo's characterizations that central processing involves careful consideration of the message arguments, which may result in a positive evaluation of the arguments ("cogent and compelling") or a negative evaluation ("weak and specious"). A positive argument evaluation, in turn, leads to favorable thoughts in response to the communication, whereas a negative argument evaluation leads to unfavorable responses.

Petty and Cacioppo also explain that during central processing, “the most important determinant of the nature of the cognitive responses elicited resides in the quality of the arguments presented in the communication” (1983, p. 7). They are indefinite, however, when it comes to what exactly an evaluation of argument quality involves (cf. Schellens & De Jong, 2004). So, taking the central route should involve a critical evaluation of message content, but what happens during that critical evaluation remains unclear. More research is needed to ascertain what exactly language users do when they take the central route to persuasion.

The ELM suggests that under conditions of high motivation and ability, argument quality should play an essential role in the persuasiveness of the message. The outcome of the persuasion process depends on the argument quality: strong arguments will lead to persuasion, whereas weak arguments will not. Considering the leading part argument quality is predicted to play in the central route to persuasion, one would expect a very clear notion of argument strength in the ELM. However, this appears not to be the case. It has been stated regularly that although argument quality is a basic construct in persuasion theory, there is a lack of focus on the question of what exactly characterizes a strong or a weak argument (e.g. Fishbein & Ajzen, 1981; McGuire, 2000; O’Keefe, 2002, 2008; O’Keefe & Jackson, 1995; Petty & Cacioppo, 1986a, 1986b). Despite several calls to pay more attention to argument quality, the nature of argument quality still lacks sufficient clarification.

The way in which the concept of argument quality is operationalized in the ELM has been a frequent topic of discussion and object of criticism (e.g. Areni, 2003; Areni & Lutz, 1988; Hornikx, 2005a; O’Keefe, 1995, 2002; Van Enschot-Van Dijk, Hustinx & Hoeken, 2003). In the ELM, argument quality has been defined in an empirical way. First, the researchers themselves developed arguments that they intuitively regarded as either strong or weak. Second, a pre-test was conducted to assess the extent to which naive participants shared this distinction, by having them rate these arguments for persuasiveness. Third, the arguments that were perceived as strong (or weak) by both the researchers and the participants were entered into a second pre-test. Respondents were instructed to think about messages containing arguments rated as either high or low and to record their thoughts during processing. A ‘strong argument’ message was defined as one that elicited predominantly favorable thoughts during careful thinking, whereas a ‘weak argument’ message was one that yielded predominantly

unfavorable thoughts under such conditions. When the ‘strong argument’ and ‘weak argument’ messages were presented to another group of respondents in a subsequent experiment, strong arguments were generally found to be more effective than the weak arguments under conditions of high elaboration.

Operationalizing argument quality in terms of its observed persuasive effects makes it impossible to test ELM-predictions concerning the relationship between argument quality and persuasive effects. Under conditions of high elaboration, people are motivated and able to carefully examine the arguments in the message. The ELM predicts that if the arguments are considered weak, unfavorable thoughts are likely to be evoked and hence, the message will probably be ineffective. If the arguments are considered strong, however, they will probably lead to favorable thoughts, after which persuasive success is more likely to occur (O’Keefe, 2008). If a message that is supposed to be a strong argument message would not lead to greater persuasion under conditions of high elaboration, the conclusion would not be that ‘the ELM-prediction has been disconfirmed’. Instead, the conclusion would be that ‘the manipulation of argument quality or amount of elaboration was inadequate’, as by *definition*, stronger arguments lead to greater persuasion when people are scrutinizing them (O’Keefe, 2002). Therefore, this called for a conceptually meaningful definition of argument quality and better insight into the specific characteristics that determine argument strength, in order to develop standards to manipulate argument quality in a way that does not depend on the judgments of participants (see, e.g. Areni, 2003; O’Keefe, 1995, 2002, 2008; Petty & Cacioppo, 1986a).

Not only does the ELM predict a connection between argument quality and persuasive outcomes, the model also predicts what factors influence the effect variations in argument quality have on persuasiveness. Whether or not argument quality matters may be dependent on such factors as topic involvement, need for cognition, and the presence of distraction or relevant background knowledge. A variety of other factors are believed to play a role in the relationship between argument quality and persuasiveness, such as the mood of the receiver and other emotional states (Petty, Fabrigar & Wegener, 2003; Wegener & Petty, 1996)<sup>4</sup>. To be able to test the validity of predictions regarding (modifying) persuasion variables, the quality of arguments should be operationalized in an adequate way.

---

<sup>4</sup> For a review of various persuasion variables, I refer to Petty (1994), Petty and Wegener (1998), and Petty, Wegener and Fabrigar (1997).

Differences in argument quality also play an important methodological role in persuasion research, as they can be used to assess whether or not respondents have been scrutinizing the message. Only participants who pay close attention to the arguments will notice the difference between messages containing strong or weak arguments. Therefore, if respondents reject the message claim supported by weak arguments or accept the claim supported by strong ones, they have been scrutinizing the message. Respondents who do not respond differently to messages containing strong or weak arguments are believed to have processed the message more superficially. So, if argument quality is not adequately operationalized, it cannot be determined whether or not respondents have been processing the message carefully (cf. Van Enschot-Van Dijk, Hustinx & Hoeken, 2003, p. 320).

## **1.2 Argument quality in argumentation studies**

According to Van Eemeren, Grootendorst and Snoeck Henkemans (1996), the study of argumentation deals with the problems involved in the production, analysis, and evaluation of argumentative discourse in light of the actual circumstances in which the argumentative discourse takes place. Broadly speaking, four problem areas in the study of argumentation can be identified: unexpressed elements in argumentative discourse, argumentation structures, argumentation schemes and fallacies (Van Eemeren, Grootendorst & Snoeck Henkemans, 1996, p. 12).

Argument quality is directly related to two of the four problem areas mentioned above. First, the quality of arguments is relevant to those who are interested in argumentation schemes, which are “conventionalized ways of displaying a relation between that which is stated in the explicit premise and that which is stated in the standpoint” (Van Eemeren, Grootendorst & Snoeck Henkemans, 1996, p. 20). In argumentation theory, different argumentation schemes have been distinguished. It is also determined under what conditions each scheme should be acceptable. Specific critical questions<sup>5</sup> have been formulated to evaluate the acceptability of an argumentation scheme, such as (for practical, goal-oriented reasoning): “Are there other ways of realizing the goal *G*?” or “Is it possible for an agent *a* to do action *A*?” (Walton, 1996, p. 11).

---

<sup>5</sup> For a discussion of argumentation schemes and critical questions, see Godden and Walton (2007).



The second area of interest related to argument quality is the study of fallacies. Many different definitions of fallacy have arisen in argumentation literature. According to Van Eemeren, Grootendorst and Snoeck Henkemans (1996), however, "It is now generally conceded that patterns of argument once considered uniformly fallacies are, in fact, fallacious in only some cases" (p. 181). That means that some patterns of argument that were traditionally considered to be fallacies, such as the *argumentum ad hominem* or the *argumentum ad verecundiam*, may in fact be acceptable variants depending on whether they meet certain criteria (Tindale, 2007). Argumentation scholars are concerned with the analysis of various kinds of fallacies and with the development of criteria to distinguish fallacious argumentation from acceptable argumentation.

From these areas of interest, it appears that the study of argumentation is concerned with the evaluation of argument quality. More specifically, it involves developing sets of criteria to distinguish normatively strong arguments from normatively weak arguments. These criteria can be more general in nature, such as the criteria of *relevance*, *sufficiency* and *acceptability* (Van Eemeren, Grootendorst & Snoeck Henkemans, 1996, p. 178), but they can also be specific to a certain argument scheme. These scheme-specific criteria are often formulated in the form of critical questions.

Argumentation studies may contribute to a further understanding of argument quality in persuasion research, as standards have been formulated for strong arguments, independent of observed persuasive effects. If an argument meets the standards, it should be considered strong. If an argument fails to meet these standards, it should be considered weak. This normative approach to argumentation can be useful, since "once one has an independently-motivated account of argument quality, it is possible to undertake empirical work that directly explores the relationship of argument quality to persuasive effects" (O'Keefe, 1995, p. 14).

Second, argumentation theory can specify what has been left vague in the dual-processing theories: the process of evaluating message content. As Schellens and De Jong (2004) put it:

"normative theories on the quality of argumentation enable us to formulate hypotheses about the central processing of persuasive texts. The ideal central processor of argumentation in persuasive texts operates in accordance with what normative theories about the quality of arguments prescribe" (p. 298).

Normative theories of argumentation prescribe that judges of argumentation should assess the extent to which a specific argument meets the criteria for being strong, criteria that may differ between different patterns of argument, or argument schemes. In this way, normative argumentation theory can answer the question of what *should* happen when people evaluate arguments carefully.

In sum, normative approaches to argumentation provide a valuable framework for studying argument quality in the ELM. After all, they allow us to form expectations on argument evaluation during central processing and enable us to empirically study the relationship between argument quality and persuasive effects.

The study of argumentation can be normative as well as descriptive. Descriptive approaches to argumentation intend to describe what *is* regarded as valid, whereas normative approaches to argumentation intend to determine what *should be* regarded as valid. The most essential difference between both kinds of approaches is that the descriptive approach is more empirically-based, whereas the normative approach has an analytical basis (Van Eemeren, Grootendorst & Snoeck Henkemans, 1996, p. 24). It has been argued that these two approaches should be closely interrelated (e.g. O'Keefe, 1995, Schellens & De Jong, 2004; Van Eemeren & Grootendorst, 2004; Van Eemeren & Houtlosser, 2007). The rationale for this belief is generally as follows: observations of argumentative behavior can be used to start thinking about norms that argumentation should meet. On the other hand, argumentation-theoretical norms are usually intended for ordinary language users, which means that these norms have more relevance if they describe argumentative conduct well. A descriptive, empirically-based approach is needed to see how well norms that should govern argumentation describe individual behaviour.

One way in which the normative approach and the descriptive approach to argumentation can be brought closer together is to determine whether norms that have been suggested on an analytical basis correspond to the norms used by ordinary language users during argument evaluation. The agreement between theoretical norms and laymen norms, however, has rarely been the object of investigation in the field of argumentation. An exception is the research by Van Eemeren and his associates (for an overview, see Van Eemeren, Garssen & Meuffels, 2007). They carried out the project *Conceptions of Reasonableness* to find out to what extent ordinary language users find discussion moves that are considered

fallacious from a pragma-dialectical<sup>6</sup> perspective, reasonable or unreasonable. The underlying principle of these empirical studies is that rules for critical discussion should not only be effective in resolving a difference of opinion or 'problem valid', but that the rules must also be intersubjectively approved or 'conventionally valid'.

A series of experiments were conducted on several fallacies committed in different stages of a critical discussion. For instance, to test the conventional validity of the argumentation scheme rule<sup>7</sup>, the fallacies *argumentum ad consequentiam*, *argumentum ad populum*, *slippery slope* and *false analogy* were investigated, all being specific rule-violations. The overall results of these studies show that "discussion moves in which a pragma-dialectical discussion rule has been violated are consistently considered unreasonable, whereas moves in which that concerning pragma-dialectical rule has not been violated, are considered reasonable" (2007, p. 263, translated from Dutch by me, E. Š.). So, positive evidence was found for the conventional validity of the pragma-dialectical discussion rules.

Although positive evidence was found for the conventional validity of the pragma-dialectical discussion rules, additional research into laymen norms of reasonableness is useful. More specifically: more insight is needed into the ideas that laymen have about the use, correct or incorrect, of argumentation schemes. In pragma-dialectical terms, this concerns the conventional validity of the argument scheme rule. In general terms, it concerns the question to what extent the normative-theoretical evaluation questions that have been suggested for certain argument schemes play a role in laymen's argumentative competence. So far, this question has not been answered adequately.

First, to test the conventional validity of the argumentation scheme rule, the specific fallacies *argumentum ad consequentiam*, *argumentum ad populum*, *slippery slope* and *false analogy* were investigated. Although each fallacy results from an incorrect application of a certain argument scheme (either causal argumentation, symptomatic argumentation or comparison argumentation), there are more incorrect applications of each scheme

---

<sup>6</sup> The pragma-dialectical approach to argumentation aims to integrate the study of critical exchanges, called *dialectics*, with the study of language use in actual communication, called *pragmatics*. Thus, in this perspective, a dialectical view of argumentative reasonableness is combined with a pragmatic view of the verbal moves made in argumentative discourse (Van Eemeren & Houtlosser, 2006).

<sup>7</sup> The argument scheme rule is: "Standpoints may not be regarded conclusively defended if the defense does not take place by means of appropriate argument schemes that are applied correctly" (Van Eemeren & Houtlosser, 2006, p. 17).

that do not correspond to traditional fallacies. It would be useful to carry out a study that would take as a starting point the argument scheme instead of the fallacy, and that would also try to get a grip on what laymen consider to be an incorrect use of such a scheme – regardless of whether the incorrect use results in a traditional fallacy or not<sup>8</sup>.

Second, getting a grip on laymen criteria for evaluating the soundness of a certain argument scheme presupposes an adequate method to retrieve them. In the *Reasonableness*-experiments, the question “Indicate briefly why you consider the [last] discussion move reasonable or unreasonable” (2007, p. 264, translated from Dutch by me, E. Š.) appeared to be a difficult task: the respondents were hardly able to explain in explicit terms what principles they rely on when giving their reasonableness judgments. As Van Eemeren et al. conclude, “time after time it has been found that [...] for our subjects between the ages of 16 and 18, responding to the question of *why* a discussion move is reasonable or unreasonable is an extraordinarily difficult abstract matter – a matter, really, our subjects apparently have never given any thought in an explicit sense” (2007, pp. 263-264, translated from Dutch by me, E. Š.). If thoughts about norms for reasonableness do not come naturally, perhaps subjects need ‘food for thought’. Although the material and instruments in the reasonableness experiments have been proven to be sufficient for generating reasonableness judgments, they might have been insufficient for retrieving *standards* of reasonableness (assuming that such standards are actually formed).

In sum, there is a need for additional research on the criteria that laymen use for judging particular argument schemes on their reasonableness and also for a systematic comparison between laymen criteria and theoretical criteria. In addition, an adequate method needs to be developed to retrieve laymen criteria.

A combination of a normative and a descriptive approach to argumentation might be considered the best way to study argument quality. In this dissertation, I will examine if indeed arguments that conform to criteria for argument quality are more persuasive than

---

<sup>8</sup> From a pragma-dialectical perspective, such a study not only would give (another) indication of the conventional validity of the argumentation scheme rule, but it would also give more insight into the empirical reality of the intersubjective testing procedure, a procedure for checking the acceptability of an argument scheme by “determining how to scrutinize the contents of the step from the proposition that is expressed in the argumentation to the proposition that is expressed in the standpoint” (Van Eemeren & Grootendorst, 2003, p. 378).

arguments that do not conform to these criteria. If there appears to be a relationship between argument strength and persuasiveness under conditions in which people scrutinize the arguments, one can tell (a) what it was exactly that made the arguments strong or weak, that is, which normative criteria these arguments did or did not meet, and (b) if these normative criteria matter to ordinary language users.

### 1.3 Pragmatic arguments: an introduction

The outcome of a persuasive attempt may depend on the type of argument that is used. This appears from ongoing research on the relative persuasiveness of different types of argument (Hoeken, 2001a, 2001b; Hoeken & Hustinx, 2003, 2006, 2009; Hornikx, 2005b). These studies show that language users are sensitive to differences in argument type when they evaluate arguments. Therefore, it is necessary to investigate for different argument types what specific characteristics make them strong or weak. In the research program *The quality of pragmatic arguments*<sup>9</sup>, to which this dissertation aims to contribute, the focus is on a type of argument called 'pragmatic argument'. In this section, I will explain why.

It is common in daily argumentative practice that standpoints are presented in which certain acts are recommended or advised against. Examples of such standpoints are:

- (1) 17-year olds should be allowed to obtain a driver's license.
- (2) The Chinese language should be taught in secondary school.
- (3) Lying in job interviews is intolerable.
- (4) The referendum as a method of electing a mayor should be abolished.  
(Retrieved from a Dutch newspaper's internet discussion:  
<http://weblogs.nrc.nl/weblog/discussie/>, 5th November, 2007, translated from Dutch by me, E. Š.)

A frequently used type of argumentation to support standpoints in which a certain act is positively or negatively evaluated is the *pragmatic argument*<sup>10</sup>. In the positive variant of the pragmatic

---

<sup>9</sup> *The quality of pragmatic arguments* is a NWO-funded research program. The project *Argument quality and probability claims*, resulting in this dissertation, is one part of the program. The project *Argument quality and desirability claims* forms another part of the program (Timmers, to appear). A monograph on the quality of pragmatic arguments, integrating the results of both projects, is also expected to appear.

<sup>10</sup> The *pragmatic argument* is also known as the argument from consequences (Godin, 1999; Walton, 1996) and the argument based on advantages and disadvantages (Schellens & Verhoeven, 1994).

argument, the acceptability of an act is defended by referring to its positive effects. In the negative variant, the unacceptability of an act is defended by referring to its negative effects. For instance, standpoints (1) en (2) can be supported by the positive variant and (3) and (4) can be backed up by the negative variant:

- (5) 17-year olds should be allowed to pass their driving tests. At least, that is the plan according to Minister Eurlings of Transport and Public Works. He wants to restrict young people, until the age of 18, to driving only in the presence of an experienced driver without a criminal traffic record. The idea is not so much to serve young people, but rather to improve traffic safety.
- (6) China has a very rich and interesting culture and it would do everybody some good to at least get acquainted with it. I propose: one obligatory year for everybody, followed by an optional course. By the way: a serious counterbalance for the American/European dominance on the world stage would be welcome. I think that even the unsatisfied Muslims in the Arabic countries could calm down...!
- (7) If you flat-out lie during a job interview the truth will come out sooner or later. You will instantly lose all credibility.
- (8) I am against mayor referenda (and presidential elections). It only leads to undesired American-style situations. Moreover, the mayor (executive authority) is given a direct mandate by citizens, which, in my opinion, appears to damage the mandate of the municipal council (controlling authority).  
(Retrieved from a Dutch newspaper's internet discussion: <http://weblogs.nrc.nl/weblog/discussie/>, 5th November, 2007, translated from Dutch by me, E. Š.)

Pragmatic arguments can be found in several contexts, such as advertising (Schellens & Verhoeven, 1994), public information campaigns (Schellens & De Jong, 2004), policy documents (Schellens & Lagerwerf, 2003) and legal decision-making (Feteris, 2002). Thus, the pragmatic argument is a frequently occurring type of argumentation in various discursive fields.

Not only is the pragmatic argument common in various areas of argumentation, it is also known as a popular type of argument in experimental studies on the persuasion process (see, e.g., Petty and Cacioppo, 1986a, pp. 54-59). Romero (1999) observed that "Persuasion researchers typically use pragmatic arguments in persuasive messages [...] Argumentation in persuasive messages used in research point to the positive consequences of a policy change, or the positive attributes of a particular product" (p. 164).

In sum, pragmatic arguments are frequently used in ordinary argumentative practice as well as in studies on persuasion processes. Therefore, the pragmatic argument is worthy of the researcher's attention.

#### 1.4 Dimensions of a pragmatic argument: probability and desirability of consequences

To further clarify the nature of the pragmatic argument, I will present some characterizations of the argument type that can be found in argumentation literature. Perelman and Olbrechts-Tyteca regard as a pragmatic argument an argument “which permits the evaluation of an act or an event in terms of its favorable or unfavorable consequences” (1969, p. 266). They perceive the pragmatic argument as a type of argument based on a cause-and-effect relation, which can only be effective if the causal link and the value of the consequence are apparent to the audience (cf. Van Eemeren, Grootendorst & Snoeck Henkemans, 1996, pp. 111-112).

In line with Perelman and Olbrechts-Tyteca, Schellens (1985; 1986) describes pragmatic argumentation as “argumentation for or against an intended action on the basis of the desirability or undesirability of its effects” and explains that pragmatic argumentation appeals to causal regularity “in the prediction of effects of the intended action” and to rules of value: “in the positive and negative valuation of effects, and in balancing these advantages and disadvantages against each other” (1986, p. 40). Schellens (1985, pp. 174-175, translated from Dutch by me, E. Š.) presents a standard scheme for the pragmatic argument:

Positive variant:	Action A leads to B B is desirable Therefore: A is desirable
Negative variant:	Action A leads to B B is undesirable Therefore: Action A is undesirable

Walton (1996) does not use the term pragmatic argument for the type of argument that Perelman and Olbrechts-Tyteca and Schellens refer to, but uses the label ‘argument from consequences’. He states that the argument from consequences is “a species of practical reasoning where a contemplated policy or course of action is positively supported by citing the good consequences of it” (p. 75). Walton (1996) also mentions the negative variant: “In the negative form, a contemplated action is rejected on the grounds that it will have bad consequences” (p. 75). He presents an argumentation scheme (1996, p. 76) that slightly deviates from Schellens’s (1985) argument scheme:

If A is brought about, then good (bad) consequences will (may plausibly) occur.  
Therefore, A should (not) be brought about.

An obvious difference between Walton's (1996) scheme and Schellens's (1985) scheme is that in Walton's scheme, the value of the consequence is included in the premise expressing a causal link between action and consequence, whereas in Schellens's scheme, the value of the consequence is expressed in a separate premise. However, it can be said that both schemes comprise probability and value.

Feteris (2002) considers various approaches to the pragmatic argument<sup>11</sup> and on the basis of these ideas comes to the following definition: "Pragmatic argumentation can be considered as a specific form of practical argumentation, of argumentation which occurs in a practical discussion about the desirability or undesirability of a certain course in action" (p. 353). With respect to the nature of the standpoint, she notes that it is a normative expression that may involve a course of action, a proposal, or a plan (involving a policy in a political context or a decision in a legal context). Regarding the nature of the argumentation, she notes that it involves the consequences of the proposed course of action. Feteris (2002, p. 355) presents a basic structure of the pragmatic argument:

Positive variant:	Standpoint: Act X is desirable Because: Act X leads to consequence Y and: Consequence Y is desirable
Negative variant:	Standpoint: Act X' is undesirable Because: Act X' leads to consequence Y' and: Consequence Y' is undesirable <sup>12</sup>

Clearly, Feteris's (2002) scheme resembles Schellens's (1985) scheme, as value (Consequence Y/Y' is desirable/undesirable) and probability (Act X/X' leads to consequence Y/Y') are expressed in separate premises.

The various characterizations of the pragmatic argument that can be found in argumentation literature show that the argument is

---

<sup>11</sup> Feteris (2002) considers argumentation theorists' perspectives as well as legal theory approaches.

<sup>12</sup> In Feteris (2002), it actually says "Consequence Y' is desirable" (p. 355). It is likely, however, that a printing error was made and that it should have said: Consequence Y' is undesirable. After all, the model is formulated for the negative variant of pragmatic argumentation.



divided into two parts: a descriptive statement saying that a certain act leads to a certain consequence and a normative statement expressing that this consequence is desirable or undesirable. The former statement expresses the consequence's *probability*; the latter statement the consequence's *desirability*. This has also been observed by different scholars concerned with persuasion research (e.g. Areni & Lutz, 1988; Hoeken, 1997, 2001a, 2001b; Hornikx, 2005a).

### 1.5 The quality of pragmatic arguments

For a pragmatic argument to be strong, the consequence's probability and the consequence's desirability or undesirability should be accepted by the reader or listener. This has been suggested in normative argumentation theory for the evaluation of the pragmatic argument. Critical questions have been formulated to determine whether the proposed course of action does indeed lead to the results and whether these results are indeed desirable (or undesirable) (e.g. Feteris, 2002; Schellens & Verhoeven, 1994). Other critical questions for the pragmatic argument have also been suggested, such as whether the proposed action is feasible or acceptable (Schellens, 1985) or whether any other (positive/negative) consequences should be taken into account (Walton, 1996). However, the questions of probability and desirability appear to be most important (cf. Feteris, 2002, pp. 357-358; Hornikx, 2005a, p. 25).

Empirical persuasion research, too, postulates that desirability and probability determine the quality of the pragmatic argument. Hoeken (2001a), for instance, states that those who put forward a pragmatic argument to persuade others "claim that their option will probably or certainly result in desirable consequences" (p. 425). Then he goes on: "The strength of their argument depends on two aspects: The consequence's desirability and the consequence's probability. A strong argument in favor of the option would be that the option will certainly result in desirable consequences" (p. 425).

Thus, the quality of the pragmatic argument should be dependent on the acceptability of the two different statements comprising the pragmatic argument. This is, however, not a sufficient answer to the question of what characteristics determine the quality of a pragmatic argument. After all, one could ask: under what circumstances should the separate statements be accepted? The evident answer is: that depends on the supporting sub-argumentation

that may be brought forward to make the statements more acceptable. This line of reasoning is represented in Figure 1.1<sup>13</sup>:

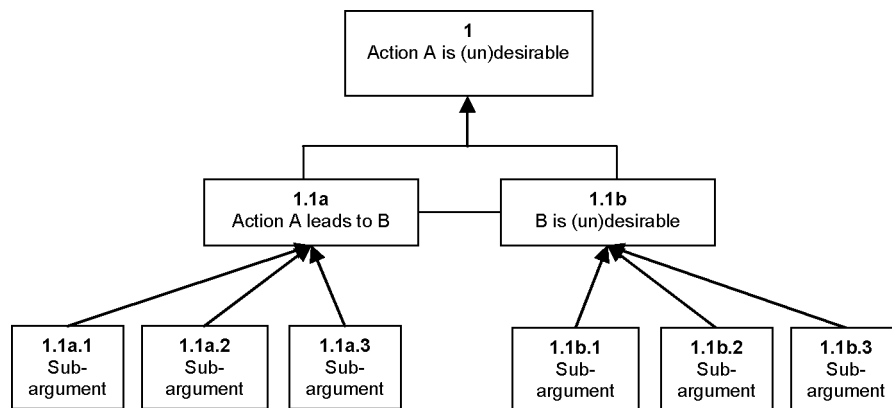


Figure 1.1 Pragmatic argument and sub-arguments supporting separate statements.

The figure shows that jointly, statements 1.1a and 1.1b can be considered sufficient to support standpoint 1<sup>14</sup>. Each element of the argument can be supported by additional arguments (In Figure 1.1, I simply chose to represent three additional arguments, but the number of additional arguments may, of course, vary). Hence, the quality of the pragmatic argument (made up of 1.1a and 1.1b) is dependent on the quality of sub-arguments supporting statement 1.1a and statement 1.1b. In other words, to understand what characteristics determine the quality of a pragmatic argument, it is necessary to understand what characteristics determine the quality of an argument in support of a probability claim and what characteristics determine the quality of an argument in support of a desirability claim. This dissertation will address the first issue. In terms of the elements in Figure 1.1: this dissertation will be concerned with the relationship between the probability-expressing statement ‘Action A leads to B’ (1.1a) and its supporting sub-arguments (1.1a.1, 1.1a.2, 1.1a.3). The other part of the model, regarding the connection between the desirability-expressing statement (1.1b) and the supporting sub-arguments (1.1b.1, 1.1b.2, 1.1b.3), is treated in another dissertation (Timmers, to appear).

<sup>13</sup> The figure is composed according to the pragma-dialectical way of representing the structure of argumentation (Van Eemeren & Snoeck Henckemans, 2006, pp. 62-65).

<sup>14</sup> In practice however, statement 1.1b is sometimes left implicit and the conclusion 1 unexpressed (Schellens & De Jong, 2004).

## 1.6 Types of arguments supporting the probability statement

Different types of argument can be brought forward in making the probability statement (more) acceptable. Based on corpus-analytical research it appears that writers do indeed support pragmatic arguments with subarguments (e.g. Hornikx, Starren & Hoeken, 2003; Schellens & De Jong, 2004). The factors determining the quality of supporting arguments will be further investigated. I chose three different argument types for further investigation. Two criteria guided this selection.

The first criterion was: the argument types should be capable of supporting the probability statement as part of a pragmatic argument. That means that the argument type should be able to result in a conclusion in the form of 'Action A leads to B'. In a pragmatic argument, this prediction is in fact a conditional sentence: *if* the action is carried out, *then* a certain consequence might be expected (cf. Schellens, 1985, p. 92). After all, action A is proposed and has not been executed yet. So, argument types were selected that could support such a conclusion.

The second criterion was: the argument types should be relatively uncontroversial, in the sense that argumentation scholars should not strongly disagree on the identification of these types. Thus, argument types were selected that are generally accepted in argumentation theory.

Three argument types meet these requirements: the argument from authority, the argument from cause to effect and the argument from example. The argument from authority is used to increase the acceptability of a conclusion by appealing to a certain source that is presumed to agree with that conclusion; the argument from cause to effect is used to predict that if something occurs, it will lead to something else; the argument from example reasons from one or more individual cases to a general statement.

Each type of argument meets the two requirements mentioned above. First, they all have the potential to support a statement of the type 'Action A leads to B' that is a part of a pragmatic argument (see Figure 1.2). One may appeal to a source saying that action A leads to B (argument from authority). The prediction that action A leads to B may also be inferred from either a more general causal relationship or from two or more predictions (argument from cause to effect). Last, the statement that action A leads to B may be supported by individual cases exhibiting the causal relationship in the statement (argument from example).

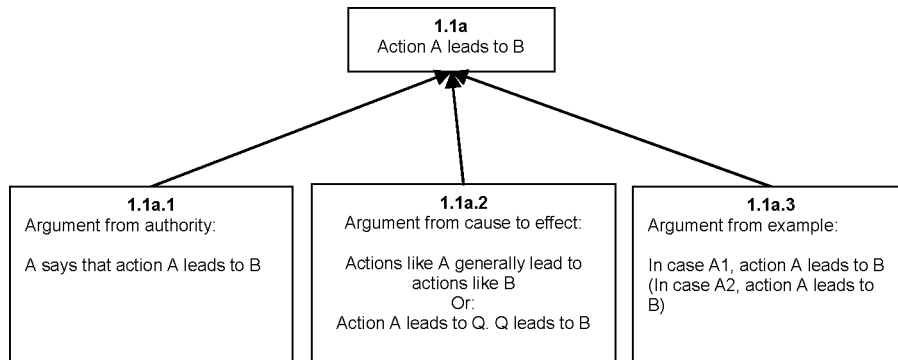


Figure 1.2 Probability claim supported by the argument from authority, the argument from cause to effect and the argument from example.

Second, the argument types appear to be relatively uncontroversial in argumentation theory. This can be inferred from how many times these argument types have been identified in argumentation theory. Garssen (1997) studied seven typologies of argumentation schemes from different perspectives and compared these with the pragma-dialectical typology. He found that argumentation from example/generalization was mentioned in all eight different typologies (Ehninger & Brockriede, 1963/1978; Freeley, 1993; Hastings, 1962; McBurney & Mills, 1964; Perelman & Olbrechts-Tyteca, 1969; Kienpointner, 1992; Schellens, 1985; Van Eemeren & Grootendorst, 1992). According to Garssen (1997), argumentation from authority and argumentation from cause to effect were mentioned in six of the eight typologies. Both types were not distinguished as a separate argument type by McBurney and Mills (1964) and Freeley (1993)<sup>15</sup>. Hence, the argument from authority, the argument from cause to effect and the argument from example often appear to be identified in argumentation theory.

<sup>15</sup> For the argument from cause to effect, I rely on Garssen (1997) suggesting that the argument from cause to effect has been distinguished as a separate type of argument by Ehninger and Brockriede (1963/1978) (see p. 31, 37) and *not* by McBurney and Mills (1964) (see p. 28, 37). This, however, does not correspond to the overview he presents at the end of his treatise, suggesting that it is the other way around (1997, p. 121). Furthermore, it should be noted that although Freeley (1993) and McBurney and Mills (1964) do not mention the argument from cause to effect as a separate type of argument, they consider causal argumentation to be a separate category that covers both the argument from cause to effect and the argument from effect to cause.

## 1.7 Aim of research and research questions

This dissertation is concerned with the quality of arguments in support of a probability claim. Argument types that may increase the acceptability of a probability claim and that are selected for further examination are the argument from authority, the argument from cause to effect and the argument from example. This leads to the following research aim:

Aim of research: To provide insight into the specific characteristics that determine the quality of the argument from authority, the argument from cause to effect and the argument from example in support of a probability claim.

A combination of a normative and descriptive approach to argumentation provides a valuable framework for studying argument quality in the ELM. As a contribution to persuasion research as well as to the study of argumentation, it should be investigated to what extent the criteria laymen use for judging a particular argument scheme correspond to normative-theoretical criteria. This leads to the following research questions:

Research question 1: To what extent do laymen criteria for the argument from authority, the argument from cause to effect and the argument from example correspond to those formulated in argumentation theory?

Research question 1a: What criteria have been formulated in argumentation theory to evaluate the quality of the argument from authority, the argument from cause to effect and the argument from example?

Research question 1b: What criteria do laymen use to evaluate the quality of the argument from authority, the argument from cause to effect and the argument from example?

According to the ELM, normatively strong arguments should be more persuasive than normatively weak arguments in conditions in which messages are processed centrally. However, there may be a difference between what people believe to be a strong argument and what actually persuades them. People might have formed certain ideas about what should persuade them or others, but these ideas might appear to be wrong in actuality (cf. O'Keefe, 1993). Several empirical studies have shown a low correspondence between

perceived or expected persuasiveness<sup>16</sup> on the one hand and actual persuasiveness on the other hand (for overviews, see Hoeken, 2001a, p. 430; O’Keefe, 1993, pp. 229-230). Hoeken (2001a), for instance, demonstrated a discrepancy between the perceived persuasiveness and actual persuasiveness of certain arguments. Dillard, Weber and Vail (2007a), on the other hand, concluded from a meta-analysis of studies measuring perceived and actual persuasiveness<sup>17</sup> of persuasive messages: the relationship between the two variables is positive and substantial. So, the findings regarding the relation between what should be persuasive and what actually persuades people are ambiguous. Therefore, it is possible that arguments that do conform to criteria for argument quality are in fact as persuasive as those that do not conform to these criteria. This leads to the following research question:

Research question 2: Are normatively strong arguments from authority, arguments from cause to effect and arguments from example actually more persuasive than normatively weak arguments?

By answering these research questions, I expect to shed more light on the specific characteristics that determine the quality of arguments in support of a probability claim.

## 1.8 Overview of the study

This dissertation consists of two different parts. In the first part, I focus on criteria for argument quality and answer research questions 1, 1a and 1b. Chapter 2 deals with criteria used to evaluate arguments from authority, chapter 3 with criteria to evaluate

---

<sup>16</sup> Hornikx (2008) explains that perceived persuasiveness should not be confused with expected persuasiveness, because both kinds of persuasiveness differ in the people at which this message is targeted. Perceived persuasiveness is receiver-oriented: “In fact, it is a message recipient’s judgment about how much a message (or argument) persuaded him or her”. By contrast, expected persuasiveness is sender-oriented: “How persuasive does a message sender (e.g., a participant in a study) think an argument is for another person” (pp. 557-558).

<sup>17</sup> In fact, Dillard et al. (2007a) refer to perceived and actual persuasiveness by perceived and actual effectiveness. There is no reason to believe, however, that they aim at different factors, considering their definition of perceived effectiveness: “we inferred a conceptual definition of PE [perceived effectiveness] as an estimate of the degree to which a persuasive message will be favorably evaluated—in terms of its persuasive potential—by recipients of that message” (p. 617).

arguments from cause to effect, and chapter 4 focuses on criteria to evaluate arguments from example. In each chapter, I examine criteria mentioned in argumentation literature, criteria mentioned by laymen and also compare both perspectives.

In the second part of the dissertation, I address research question 2, concentrating on the relationship between argument strength and actual persuasiveness. Chapter 5 reports on a preliminary experimental study on the relation between perceived reasonableness and actual persuasiveness of normatively strong and weak arguments<sup>18</sup>. In chapter 6, I report on an experiment on the effect of meeting certain criteria for argument quality on actual persuasiveness. Chapter 7 presents a general conclusion and a discussion, in which I reflect on the studies described in both parts of the dissertation.

The two parts of which this dissertation consists differ in three important respects: in *topic*, in *kind of argumentation research*<sup>19</sup>, and in *method of research*. First of all, it can be said that the first part is all about norms: norms that, according to argumentation theory, lay people *should* use for argument evaluation and norms that lay people *do* use for argument evaluation. The second part is all about the role that these norms might play in the persuasion process. Second of all, it can be said that the first part reports 'concept research', whereas the second part reports 'persuasion research' (cf. Van Eemeren & Garssen, 2007). The first part deals with differences between theoretical conceptions of argument quality and conceptions of lay people. The second part deals with studies on the influence of message content characteristics. Third of all, part I and II differ in a methodological respect: in the first part the research is primarily qualitative in nature, whereas the second part reports on quantitative research.

---

<sup>18</sup> This study has been published earlier (Timmers, Šorm, Schellens & Hoeken, 2008). This publication is the result of an intensive collaboration between the first and the second author who contributed in equal measure to the design, execution, and reporting of the study.

<sup>19</sup> See Van Eemeren and Garssen (2007, pp. 193-195), who distinguish between several kinds of empirical argumentation research.

## **PART I: CRITERIA FOR ARGUMENT QUALITY**





## 2. The quality of the argument from authority

This chapter deals with the quality of the argument from authority. First, in section 2.1, I will discuss different types of authority and argue that the epistemic/cognitive type of authority is relevant to further investigation. I will also discuss De George's (1985) and Wilson's (1983) view on this type of authority. They both formulated criteria that authorities should meet, but each from a different perspective: De George (1985) from a philosophical perspective and Wilson (1983) from a social epistemological point of view. Their views on authority serve as a frame of reference for the argumentation-theoretical view on the quality of the argument from authority. In section 2.2, I will address the question of what criteria are used in argumentation theory to evaluate the quality of the argument from authority and explain how these relate to the criteria formulated by De George (1985) and Wilson (1983). In section 2.3, I will address the question of what criteria are used by laymen to evaluate the quality of the argument from authority. In the final section 2.4, I will compare the different sets of criteria.

### 2.1 Introduction

#### 2.1.1 Types of authority

The concept of authority is ambiguous, as it has many different meanings depending on the circumstances. Distinctions have been made, for instance, between *epistemic authority* and *deontic authority* (Bochenski, 1974), between *de facto authority* and *de jure/iure authority* (Woods & Walton, 1974; Schellens, 1985), between *executive authority* and *nonexecutive authority* (De George, 1985) and between *expertise-based* and *power-based* authority<sup>20</sup> (Schellens & Verhoeven, 1994). Walton (1997), in line with Wilson (1983), prefers a division between two broad types of authority:

"The *cognitive (epistemic, de facto)* type of authority is a relationship between two individuals where one is an expert in a field of knowledge in such a manner that his pronouncements in this field carry a special weight of presumption for the other individual that is greater than the say-so of a layperson in that field. The cognitive type of authority, when used or appealed to in argument, is essentially an appeal to expertise, or to expert opinion. By contrast, the *administrative (deontic, de jure)* type of authority is

---

<sup>20</sup> Translated from the Dutch terms "*deskundigheid*" and "*bevoegdheid*" (Schellens & Verhoeven, 1994, p. 130).

a right to exercise command or influence, especially concerning rulings on what should be done in certain types of situations, based on an invested office, or an official or recognized position of power” (pp. 77-78).

Goodwin (1998) even identifies three types: *command*, *expertise* and *dignity* authority. Goodwin’s distinction between command authority and expertise authority corresponds to the administrative-cognitive distinction drawn by Walton (1997) and Wilson (1983), but the third authority she identifies, the dignity authority, turns out to be a different kind. Dignity authorities cannot be distinguished by knowledge or power, but rather by eminence (e.g. royal persons, priests, Nelson Mandela, or Dalai Lama). The appropriate response to such authorities is not obedience (as to command authority) or prudence (as to expertise authority), but deference - or showing respect (cf. Hansen, 2006, p. 323).

Although types of authority are classified and labelled in many different ways, the interpretations do not differ significantly. In my opinion, De George’s nonexecutive authority can be considered an umbrella term for Goodwin’s dignity and expertise authorities as they are not in the position to execute commands or control (so, they are nonexecutive authorities). For the same reason, cognitive, expertise-based, de facto and epistemic authorities belong to the category ‘nonexecutive authority’. De George’s executive authority - who is in a certain position to act for or on someone else - appears similar to Goodwin’s command authority and also to the administrative, power-based, de jure/iure, deontic types of authority. Table 2.1 shows the similarities and differences between types of authorities in the typologies mentioned above.

Table 2.1 Comparison between different typologies of authority.

<i>Typology</i>	Bochenski 1974	Woods & Walton 1974; Schellens 1985	Schellens & Verhoeven 1994	Wilson 1983; Walton 1997	Goodwin 1998	De George 1985
<i>Type of authority</i>	deontic	de jure/iure	power-based	administrative	command	executive
	epistemic	de facto	expertise-based	cognitive	expertise	non-executive
	-	-	-	-	dignity	

The next question to be answered is: what kind of authority is relevant within the framework of pragmatic argumentation? More specifically: what type of authority might be appealed to in supporting a statement that a proposed cause (probably) leads to a certain consequence? In fact, any of the distinguished kinds of authority might be used to establish such a statement. After all, all kinds of authority may express statements that are causal in nature. However, as the administrative (or deontic, *de jure/iure*, power-based etc.) and dignity kinds of authority are primarily associated with telling others what ought or ought not to be done - prescriptive statements, in other words - I consider the cognitive (or epistemic, *de facto*, expertise-based etc.) kind of authority as more relevant in support of a statement that A will (probably) lead to B. As a consequence, I will focus on this particular type of authority. Next, in section 2.1.2, I will discuss the conditions that a good epistemic/cognitive authority should satisfy according to De George (1976, 1985) and Wilson (1983).

### 2.1.2 Epistemic authority

A large number of works have dealt with authority in general, and with particular aspects of authority<sup>21</sup>. An insightful, philosophical analysis of epistemic authority is provided by De George (1976, 1985). According to his definition, "X is a *de facto* epistemic authority if there is some Y who considers X an authority for Y in some realm (R). With respect to that realm, Y considers X his superior in knowledge" (1985, p. 27). Furthermore, De George (1985) discusses the conditions under which an epistemic authority is legitimate:

"For Y's belief in what X says in R to be justified, it is necessary that: (1) X have [sic] knowledge of R (the knowledge criterion); (2) Y have [sic] good reason to believe that X has knowledge of R (the inductive criterion); (3) the *p*'s [propositions] that Y believes fall within R or be so related to R that some aspect of X's knowledge of R is sufficient to justify accepting *p* (the relevance criterion); and (4) Y have [sic] good reason for believing that X is telling the truth or stating what X believes when X enunciates *p* (the trustworthiness criterion)" (pp. 35-36).

The first criterion, the knowledge criterion, is about the question whether or not there is any knowledge for someone to possess. For

---

<sup>21</sup> For an overview, see De George's Bibliographic Essay (1985, pp. 293-299).

example<sup>22</sup>, if someone appeals to an authority specialized in the field of intelligent extraterrestrial life, the epistemic authority can be considered illegitimate because there is no such knowledge. If I interpret De George (1985) correctly, the knowledge criterion states that the epistemic authority is justified if his/her field of knowledge is a genuine field of knowledge<sup>23</sup>.

The second criterion, the inductive criterion, pertains to the question if there is sufficient reason for the audience to believe that the proposed authority is someone who has knowledge of the realm. De George (1985) acknowledges a close relation with the knowledge criterion, but explains that the inductive process “involves not only knowing that there is knowledge but also knowing who has it” (p. 49). So, the inductive criterion says that the epistemic authority is justified if the inductive basis is enough to draw the conclusion that X is knowledgeable in the field<sup>24</sup>.

The third criterion is the relevance criterion, concerning the question whether or not the statements of the proposed authority are in the realm of its competence. If they are in the realm, there is good reason to believe the proposed authority. However, “When X asserts some p in a field other than R, X’s knowledge of R is not usually sufficient to justify belief in what X says unless the second p is in a field closely related to R or is otherwise connected to it” (p. 52). So, the third standard is the relevance of the realm of competence to the field that the statement is in.

The fourth criterion is the trustworthiness criterion, referring to the question whether or not the proposed authority is a reliable source of information. According to De George (1985), this is a question of character and fairness, and “Y must have good reason for

---

<sup>22</sup> This example is not presented by De George (1985). De George (1985), however, gives a similar example: “Knowledge of how the stars determine individual destinies is simply not available if the stars do not determine individual destinies. People may claim to have such knowledge, and others may believe them. But their belief cannot be grounded if there is no such knowledge” (p. 47).

<sup>23</sup> In my opinion, the phrasing ‘X have knowledge of R’ is somewhat unfortunate. It reads as ‘The proposed authority should have knowledge of the realm’, whereas that was probably not De George’s (1985) intention. If I correctly assume that the knowledge criterion deals with the presence of real knowledge, phrasing such as ‘there is knowledge of R’ or ‘knowledge of R exists’ would have been more appropriate.

<sup>24</sup> I find the term ‘the inductive criterion’ somewhat confusing, as judging epistemic authority with respect to the other three criteria may involve inductive bases as well. There should also be, for instance, enough inductive bases to conclude that someone is trustworthy.

believing that X is telling the truth or is stating what X believes when X enunciates p" (p. 40). So, the fourth condition under which an epistemic authority is legitimate is trustworthiness.

Wilson (1983) analyzes the phenomenon of cognitive authority from the perspective of social epistemology rather than from philosophical epistemology, as De George (1985) does. He starts on the basis of the following definition:

"We shall say that person A is a cognitive authority for person B with respect to sphere of interest S to the degree that what A says about questions falling within the sphere S carries weight for B. A is a cognitive authority for me in matters of politics to the degree that what A says about political questions carries weight with me. If what A says carries a lot of weight, he has a lot of authority; if it carries no weight, he has no authority" (p. 13).

He makes three points in addition to this definition: (1) that the authority involves at least two people, because there has to be another person to be an authority for (2), that he can be one in different degrees, and (3) in different spheres. So far, his definition is very similar to De George's, although in Wilson's definition, the gradedness of cognitive authority is made explicit (expressed in "to the degree").

Wilson (1983) addresses the question of how we can choose among sources who all claim to be experts on an issue. He mentions a couple of 'bases of authority'. First of all, to answer the question 'What qualifies him to speak on the subject?', he mentions the 'occupational specialization rule' ("he makes his living dealing with that subject", p. 21) and the 'formal education rule' ("he has studied the subject systematically and deeply and has earned advanced degrees in the subject", p. 21).

Then he goes on to discuss two kinds of reputation rules to answer the question "Is he a *great* expert?": the 'reputation among peers rule' ("we take reputation among others who are supposed to be experts in the same line of work or study as indication of outstanding competence", p. 22) and the 'reputation among other cognitive authorities rule' ("Those already established as my cognitive authorities can transfer authority to another", p. 22). Wilson (1983) also mentions the 'common consent rule' that "amounts to a generalization of the two other reputation rules" (p. 23). This rule says that someone can be identified as a cognitive authority in a certain sphere, if everyone recognizes him as an authority in that sphere.

In addition, Wilson (1983) mentions the 'performance rule', stating that if a person can achieve striking results in a certain area, he or she deserves to be recognized as a cognitive authority in that

area. He notes, however, that there are many cases in which we cannot evaluate the performances ourselves, so that the performance rule cannot serve us and we have to rely on the reputation rule.

Next on Wilson's list is the 'intrinsic plausibility rule': cognitive authority can be justified on the ground that he continues to tell plausible things. He notes that prior beliefs may determine the applicability of this particular test. The more the expert talks about matters that are distant from our own established beliefs, the less likely it is that we will be able to apply the test.

The last rule is the 'personal trust rule', saying that figures acquire cognitive authority, if their personality is so impressive that we are prepared to believe whatever they say (e.g. prophets, heroes).

If we compare Wilson's bases of cognitive authority to De George's criteria for legitimate epistemic authority, two substantial differences can be observed (see Table 2.2). First of all, Wilson (1983) does not refer to the knowledge criterion and relevance criterion like De George (1985) does. Second, the inductive criterion, identified by De George (1985), appears to be further specified by Wilson (1983) into several rules. All these rules may be used to answer the question if there is sufficient reason to believe that the proposed authority is someone who has knowledge of the realm.

Table 2.2 Comparison between De George's and Wilson's approach to epistemic/cognitive authorities.

<i>De George (1985)</i>	<i>Wilson (1983)</i>
The knowledge criterion	-
The inductive criterion	Occupational specialization rule Formal education rule Reputation among peers rule Reputation among other cognitive authorities rule Common consent rule Performance rule Intrinsic plausibility rule
The relevance criterion	-
The trustworthiness criterion	Personal trust rule

In section 2.2, I will present a study on the argumentation-theoretical conceptions of good authorities and discuss how these correspond to De George's and Wilson's views.

## 2.2 Argumentation-theoretical criteria

Traditionally, the improper use of an appeal to authority has been called the fallacy of *argumentum ad verecundiam*. According to Walton (1989, 1997), logic textbooks have typically defined the fallacy as appealing to an authority of expertise. More recently, it has been acknowledged that an appeal to expert opinion is not always a fallacy, but should be regarded as a type of argument that can be reasonable in some cases but fallacious in others (Schellens, 1985; Walton, 1989, 1997). Therefore, in more recent textbooks, critical questions have been suggested to help readers determine whether or not an appeal to expertise should be considered a fallacy.

In an extensive study on appeals to expert opinion, Walton (1997) presented a survey of the sets of critical questions proposed in various textbooks (pp. 199-222). He noted: "While many accounts appear to agree on a few fundamental points, there is in fact not a great deal of uniformity or consistency in their treatments of the subject" (p. 199). On the basis of an analysis of the textbook criteria for evaluation, he formulated a new set of critical questions (pp. 222-229). Similar to Walton (1997), I made an inventory of critical questions for the evaluation of the argument from authority. However, there are some differences between his study and mine.

First, I grouped together critical questions with similar meanings, to be able to determine if these *kinds* of questions are used by laymen for evaluating the argument from authority. I did not intend to propose a new, unified set of critical questions on the basis of the analysis. In other words, the emphasis in my study is more on classification and less on identifying points of agreement and disagreement and solving problems by revising the textbook criteria.

Second, I included sets of critical questions in my literature study that Walton (1997) does not appear to have used in his. I also included the set of questions that Walton (1997) proposed.

Third, I report on the extent to which the evaluation criteria in argumentation theory correspond to the views on authority in philosophical and social epistemology (see 2.1.2).

In this section, the key question is: what evaluation questions have been suggested in argumentation theory for the evaluation of the argument from authority? To address this question, a literature study was carried out. Next, the details of the method will be discussed.



### 2.2.1 Method

A selection of publications was made in the field of argumentation theory. I only took into consideration those publications treating the argument from authority and its evaluation. To determine which argumentation-theoretical approaches would be relevant, I used Garssen (1997) as a starting point. In his work, he deals with typologies of argumentation schemes from different approaches: the pragma-dialectical approach, argument types in American debating handbooks, Hastings's typology of argument schemes, argument schemes in Perelman and Olbrecht-Tyteca's *New Rhetoric*, Schellens's typology of arguments and Kientpointer's approach. Using these as a basis, I decided to add some approaches to my literature study and to disregard some others.

First, I disregarded sources published before 1985, the year in which Schellens published his *Reasonable Arguments*. In this work, he considers existing evaluation questions associated with argument schemes and presents his own, modified version. Since Schellens had already dealt with publications before 1985, I relied on his *Reasonable Arguments* and on the work he published together with Verhoeven in 1994. In this book, Schellens and Verhoeven present a summary of occasionally adjusted evaluation questions<sup>25</sup>.

Second, I did not take into consideration the work of Perelman and Olbrecht-Tyteca, and Kientpointer. Although they distinguished different argument schemes, they have not explicitly formulated evaluation questions connected to these schemes.

Third, I decided to add two publications to the literature study that focus on *critical thinking*: Warnick and Inch (1989) and Reinard (1991). Both books aim to develop skills in constructing as well as refuting arguments. I decided to add these publications, since the perspective of critical thinking is different from other perspectives in that it is a domain not primarily defined by its subject – argumentation – but still concerned with the ability to evaluate arguments (see also Van Eemeren, Grootendorst & Snoeck Henkemans, 1996, pp. 163-188).

Fourth, I considered Meany and Shuster (2002). I selected this source, because it is more up-to-date than the debating literature used by Garssen (1997). They also mention the argument from

---

<sup>25</sup> More recently, in a chapter on argumentation analysis, Verhoeven and Schellens (2008) presented argument schemes and scheme-dependent evaluation questions. As these are based on Schellens and Verhoeven (1994), I decided to disregard Verhoeven and Schellens (2008).

authority as a separate kind of argument and formulate critical related questions.

Last, I included Walton's view on the evaluation of the argument from authority, as published in *Appeal to Expert Opinion: Arguments from Authority* (1997).

In sum, the following works were studied: Garssen (1997), Schellens (1985), Schellens and Verhoeven (1994), Walton (1997), Meany and Shuster (2002), Reinard (1991) en Warnick and Inch (1989). An inventory was made of the evaluation questions formulated in these publications for the argument from authority. Then, the evaluation questions were classified on the basis of relationships between them. Next, the results of this classification will be reported.

## 2.2.2 Results

I classified the criteria formulated in argumentation theory to evaluate arguments from authority. Eight main categories can be distinguished with questions that apply specifically to the argument from authority. I found one question that can be applied not only to the argument from authority, but also to other argument schemes. Next, I will discuss each of the categories.

### (I) Specific to the argument from authority

#### (I.1) Relevant expertise criterion

In argumentation theory, questions have been proposed about whether or not the authority cited is an expert in the relevant field, that is, the field to which the assertion pertains (see Table 2.3). This group of questions refers to what I call the *relevant expertise criterion* (I.1).

There is also a category of questions that refer to what I label the *expert criterion* (I.1.1). These questions ask whether or not the source is really an expert. In my opinion, it is impossible to determine whether or not someone's expertise is relevant to the field to which the statement pertains, before the question is answered whether the source indeed has some kind of expertise. Therefore, I consider the *expert criterion*-category (I.1.1) a subcategory falling under the more general category *relevant expertise criterion* (I.1).

Furthermore, Reinard (1991) proposes the following criterion: "Is the source respected by other authorities?"<sup>26</sup> (p. 148). As this

---

<sup>26</sup> Walton (1997) suggests a similar question: "Can testimony of peer experts in the same field be given to support *E*'s competence?" (p. 223). He suggests this

question may help to determine if someone is credible as an expert source, I created a special category labelled *respect by other authorities criterion* (I.1.1.1).

Table 2.3 Evaluation questions falling under the category of (I.1) Relevant expertise criterion.

<i>Evaluation question</i>	<i>Source</i>
I.1 Relevant expertise criterion	
Is the source well qualified to speak on the issue?	Warnick & Inch (1989, p. 114)
Is <i>E</i> an expert in the field that <i>A</i> is in?	Walton (1997, p. 223)
Is the person who is proposed as an authority actually an expert in the domain concerned? <sup>27</sup> (preliminary question)	Garsen (1997, p. 11)
Is the authority of the one who makes the statement a sign of correctness or acceptability of that statement? <sup>28</sup>	Garsen (1997, p. 11)
Are they qualified to speak about the subject they are cited in?	Meany & Shuster (2002, p. 63)
I.1.1 Expert criterion	
How credible is <i>E</i> as an expert source?	Walton (1997)
Is the source competent?	Reinard (1991, p. 147)
Is <i>A</i> an expert? <sup>29</sup>	Schellens (1985, p. 186); Schellens & Verhoeven (1994, p. 131)
What are the qualifications of the person(s) cited as a source?	Meany & Shuster (2002, p. 63)
I.1.1.1 Respect by other authorities criterion	
Is the source respected by other authorities?	Reinard (1991, p. 148)

### (I.2) Trustworthiness criterion

This category contains questions referring to the reliability of the authority (see Table 2.4). These questions refer to the extent to which the authority's *person* (as opposed to his *expertise*) is relied upon or confided in. To get an idea of the trustworthiness of the cited authority, argumentation theorists have suggested considering the factors bias and honesty. Walton (1997) explains that it is acceptable

---

question as a critical subquestion under the 'expertise question', which is similar to what I call the *expert criterion*.

<sup>27</sup> In Dutch: "Is de persoon die als autoriteit wordt opgevoerd, inderdaad deskundig op het betreffende gebied?" (1997, p. 11).

<sup>28</sup> In Dutch: "Is het gezag of de autoriteit van degene die de uitspraak doet wel een teken van de juistheid of aanvaardbaarheid van die uitspraak?" (1997, p. 11.)

<sup>29</sup> In Dutch: "Is *A* deskundig?" (1985, p. 186) and "Is de aangehaalde autoriteit deskundig genoeg om het standpunt op zijn gezag te aanvaarden?" (1994, p. 131).

for someone to have a bias, but the opinion should be based on knowledge in the field and not on a motive for gain (a “bad” bias).

Table 2.4 Evaluation questions falling under the category of (I.2) Trustworthiness criterion.

<i>Evaluation question</i>	<i>Source</i>
I.2 Trustworthiness criterion	
Is the source unbiased?	Warnick & Inch (1989, p. 114)
Is the source biased so much that the opinion is unreliable?	Reinard (1991, p. 147)
Is the source more or less biased about the topic at hand?	Meany & Shuster (2002, p. 63)
Are there any reasons to doubt A's [...] objectivity? <sup>30</sup>	Schellens & Verhoeven (1994, p. 131)
Is A sincere? <sup>31</sup>	Schellens (1985, p. 187); Schellens & Verhoeven (1994, p. 131)
Is E personally reliable as a source?	Walton (1997, p. 223)

### (I.3) Accuracy of the cited opinion criterion

This class contains questions referring to the relationship between what the cited authority actually asserted and what is presumed to be his opinion (see Table 2.5). In other words, all questions within this category ask whether the cited opinion was accurately inferred from what the individual in fact asserted. Walton (1997) calls this the *opinion question* or the *assertion question* and explains that it “relates to the opinion asserted by the expert and to how what is presumed to be his opinion (a proposition) was extracted or inferred from the text of discourse the expert actually put forward in dialogue with the respondent” (p. 225).

<sup>30</sup> In Dutch: “Zijn er redenen om te twijfelen aan A's [...] objectiviteit?” (only in Schellens & Verhoeven, 1994, p. 31, not in Schellens, 1985).

<sup>31</sup> In Dutch: “Is A oprecht?” (1985, p. 187); “Zijn er redenen om te twijfelen aan A's oprechtheid [...]?” (1994, p. 131).

Table 2.5 Evaluation questions falling under the category of (I.3) Accuracy of the cited opinion criterion.

<i>Evaluation question</i>	<i>Source</i>
I.3 Accuracy of the cited opinion criterion	
What did <i>E</i> assert that implies <i>A</i> ?	Walton (1997, p. 223)
Has the cited authority indeed claimed <i>P</i> ? <sup>32</sup>	Schellens & Verhoeven (1994, p. 131)
Has <i>P</i> been put forward correctly? Has it not been taken out of context (...)? <sup>33</sup>	Schellens (1985, p. 186); Schellens & Verhoeven (1994, p. 131)
Is <i>P</i> clear and unambiguous? <sup>34</sup>	Schellens (1985, p. 186); Schellens & Verhoeven (1994, p. 131)

#### (I.4) External consistency criterion

According to argumentation theory, another way to evaluate the argument from authority is asking about the external consistency of the expert opinion (see Table 2.6). The advanced opinion may or may not be consistent with the opinion of other experts or with other available information. It can be argued that the external consistency-question should serve as a sub-question to answer other questions. For instance, from an opinion that is externally inconsistent, it can be concluded that the cited individual is not a genuine expert (because 'genuine experts say what other experts say'), or not trustworthy ('if the expert's statements are not in accordance with statements of other experts, he may not be relied upon').

<sup>32</sup> In Dutch: "Heeft de aangehaalde autoriteit inderdaad *P* beweerd?" (only included in Schellens and Verhoeven, 1994, p. 131, not in Schellens, 1985).

<sup>33</sup> In Dutch: "Is *P* juist weergegeven, niet uit zijn verband gerukt (...)?" (1985, p. 186).

<sup>34</sup> In Dutch: "Is *P* duidelijk en ondubbelzinnig?" (1985, p. 186); "Is het standpunt *P* duidelijk en ondubbelzinnig onder woorden gebracht?" (1994, p. 131).

Table 2.6 Evaluation questions falling under the category of (I.4) External consistency criterion.

<i>Evaluation question</i>	<i>Source</i>
I.4 External consistency criterion	
Is the testimony consistent with other sources of information?	Warnick & Inch (1989, p. 114)
Is the opinion externally consistent?	Reinard (1991, p. 150)
Is P in accordance with statements of other authoritative sources? <sup>35</sup>	Schellens (1985, p. 186); Schellens & Verhoeven (1994, p. 131)
Is P in accordance with other available information? <sup>36</sup>	Schellens (1985, p. 186); Schellens & Verhoeven (1994, p. 131)
Is A consistent with what other experts assert?	Walton (1997, p. 223)

#### (I.5) Internal consistency criterion

Other critical questions refer to the internal consistency of the opinion, so the extent to which the expert's statement contradicts his or her other statements (see Table 2.7). Similar to external consistency, it can be argued that the question of internal consistency should serve as a sub-question to answer other questions, like the expert-question or the trustworthiness-question. If someone appears to have changed his opinion, this may raise the question if he or she is really that competent ('competent people do not change their opinion or do not make contradictory statements') or reliable ('if he changes his opinion from time to time, he may not be relied upon').

Table 2.7 Evaluation questions falling under the category of (I.5) Internal consistency criterion.

<i>Evaluation question</i>	<i>Source</i>
I.5 Internal consistency criterion	
Is the opinion internally consistent?	Reinard (1991, p. 149)
Is P in accordance with other statements from the same source? <sup>37</sup>	Schellens (1985, p. 186); Schellens & Verhoeven (1994, p. 131)

<sup>35</sup> In Dutch: "Is P in overeenstemming met uitspraken van andere gezaghebbende bronnen?" (1985, p.186).

<sup>36</sup> In Dutch: "Is P in overeenstemming met andere beschikbare informatie?" (1985, p. 186); "Is het standpunt P in overeenstemming met [...] informatie van andere gezaghebbende bronnen?" (1994, p. 131).

<sup>37</sup> In Dutch: "Is P in overeenstemming met andere uitspraken van dezelfde bron?" (1985, p. 186).

### (I.6) Ability to provide evidence criterion

Walton (1997) recommends asking for backup evidence (see Table 2.8). If the expert says something about a particular situation that cannot be verified by the receiver, then evidence is needed to understand how the expert came to have this opinion. If the proposed authority does not react appropriately to this request for additional information, then his or her opinion is less likely to be believed. A similar question has been suggested by Warnick and Inch (1989): "Has the source access to firsthand information about the issue?" (p. 114). This question is comparable to Walton's, as the expert should back up the statement with evidence that he or she personally had access to the situation or information pertaining to the statement. Again, it can be argued that the evidence test should serve to answer other questions. For instance, if the cited authority does not provide evidence on which the opinion was based, this authority may be valued as 'low' in the expert-dimension ('genuine experts are able and willing to give additional information') and on trustworthiness ('if he does not comply with the request to give additional information, he may not be relied upon').

Table 2.8 Evaluation questions falling under the category of (I.6) Ability to provide evidence criterion, (I.7) Recency criterion, (I.8) Opinion-to-conclusion relevance criterion and (II) Strength of conclusion criterion.

<i>Evaluation question</i>	<i>Source</i>
I.6 Ability to provide evidence criterion	
Is A's assertion based on evidence?	Walton (1997, p. 223)
Has the source access to firsthand information about the issue?	Warnick & Inch (1989, p. 114)
I.7 Recency criterion	
Is the cited statement not outdated? <sup>38</sup>	Schellens (1985, p. 186)
I.8 Opinion-to-conclusion relevance criterion	
(...) has it [P] (possibly) been correctly interpreted or paraphrased? <sup>39</sup>	Schellens (1985, p. 186); Schellens & Verhoeven (1994, p. 131)
II Strength of conclusion criterion	
Is the certainty of the conclusion in accordance with the answers on question 1-9? <sup>40</sup>	Schellens & Verhoeven (1994, p. 132)

<sup>38</sup> In Dutch: "Is de aangehaalde uitspraak niet verouderd?" (only in Schellens, 1985, p. 186, not in Schellens & Verhoeven, 1994).

<sup>39</sup> In Dutch: "Is P (...) (eventueel) correct geïnterpreteerd of geparafraseerd?" (1985, p. 186).

<sup>40</sup> In Dutch: "Is de stelligheid van de conclusie in overeenstemming met de antwoorden op vraag 1-9?" (only in Schellens & Verhoeven, 1994, p. 132, not in Schellens, 1985).

### (1.7) Recency criterion

Schellens (1985, p. 186) included the following in his set of questions: "Is the cited statement not outdated?" (see Table 2.8). Even if an argument from authority passes all other tests (relevant expertise, trustworthiness, etc.), the receiver may still decide that the statement is outdated and not accept the supported claim. It is also possible that the recency-question serves as a sub-question to determine the degree of expertise. After all, if a statement has been made years ago, it is reasonable to think that the cited individual was, at the time he made his assertion, not as competent as those individuals who gained knowledge in the domain more recently.

### (1.8) Opinion-to-conclusion relevance criterion

In Schellens (1994) and in Schellens and Verhoeven (1994), a criterion is mentioned concerning the relevance of the source's presumed or cited opinion to the claim that is defended. I call this criterion the *opinion-to-conclusion relevance criterion*. It can be considered a scheme-specific variant of the more general relevance criterion. For example:

1. A leading Australian eye expert has warned that people who do not adopt a healthy life style can expect a loss of vision in later life. Therefore, an unhealthy lifestyle can lead to vision loss.
2. That bag of chips you're nibbling on today could affect your eyesight tomorrow ... that's the warning from a leading Sydney optometrist published in a Brisbane newspaper. Therefore, junkfood causes blindness.  
(The first example is based on <http://www.news-medical.net/news/> and the second example is based on <http://www.pca.com.au/healthnews.php>, retrieved June 9th, 2009)

In the first example, the expert's warning appears to prove the right conclusion, whereas in the second example, the warning appears to prove the wrong conclusion. Or, to avoid the right-wrong dichotomy: the conclusion in the first example is more appropriate than the conclusion in the second example. The second example appears to be an instance of a fallacy of relevance and more specifically, of a subtype that according to Walton (2004) could be correctly described as the *wrong conclusion fallacy*: "The simplest type of fallacy of relevance occurs in cases where the chain of argumentation starting from the premises proves the wrong conclusion, i.e. a proposition that may look like the conclusion to be proved, but is a different



conclusion<sup>41</sup>” (p. 72). For an argument from authority specifically, it seems reasonable to require that the presumed opinion does not leap too quickly to the conclusion it is supposed to be argued for.

(II) Generally applicable to all argument schemes: strength of conclusion criterion

As Table 2.8 shows, I found one question that I do not consider to be specific to the argument from authority: “Is the certainty of the conclusion in accordance with the answers on question 1-9?” (Schellens & Verhoeven, 1994, p. 132). After all, it goes for all argument types that the conclusion should not be formulated in terms that are too absolute.

### **2.2.3 Conclusion and discussion**

In this section, an attempt was made to answer the following question: what criteria have been formulated in argumentation theory to evaluate the quality of the argument from authority? It appears that the suggested criteria can be classified according to eight different main categories that apply specifically to the argument from authority: (I.1) the relevant expertise criterion, (I.2) the trustworthiness criterion, (I.3) the accuracy of the cited opinion criterion, (I.4) the external consistency criterion, (I.5) the internal consistency criterion, (I.6) the ability to provide evidence criterion, (I.7) the recency criterion and (I.8) the opinion-to-conclusion relevance criterion.

As I mentioned in the discussion of the various classes, it can be argued that some kinds of questions should serve as sub-questions to answer other questions. More specifically, questions in classes (I.4)-(I.7) might be used for answering questions that fall into category (I.1) and (I.2). Therefore, category (I.1) and (I.2) may be regarded as two overarching categories.

To what extent do the classes of evaluation criteria correspond to De George’s and Wilson’s view to good authorities, as discussed in 2.1.2? First, I discuss the similarities. As Table 2.9

---

<sup>41</sup> According to Walton (2004), this is the fallacy of *ignoratio elenchi*, or ignorance of refutation, described by Aristotle. He also treats another fallacy of relevance, which is known as red herring: “the fallacy of switching the topic of the discussion to some different issue that may be highly entertaining and diverting to the audience, distracting its attention from the real issue to be discussed” (p. 72).

shows, De George's inductive criterion and most of Wilson's rules correspond to the argumentation-theoretical expert criterion. I already argued that criteria in category (I.4)-(I.7) may be used to conclude that someone is a genuine expert. Therefore, it can be said that these categories correspond to De George's inductive criterion. De George's relevance criterion is similar to the relevant expertise criterion found in argumentation theory. Furthermore, De George's trustworthiness criterion and Wilson's personal trust rule correspond to the argumentation-theoretical trustworthiness criterion.

There are also differences between argumentation theory and the views of De George and Wilson. De George's knowledge criterion, saying that someone should possess genuine knowledge, does not correspond to a general class of questions in argumentation theory. That does not mean that such a criterion cannot be found at all in argumentation theory. Walton (1997) has recommended it as a subquestion under what he calls the field-question: "Is the field of expertise cited in the appeal a genuine area of knowledge, or area of technical skill that supports a claim to knowledge?" (p. 224). In my opinion, it may rather serve as a subquestion for the expert-question, because someone cannot be a genuine expert if the area of knowledge is not genuine.

Furthermore, the argumentation-theoretical questions about the accuracy of the cited opinion (I.3) and the relevance of the opinion to the conclusion (I.8) have no equivalent in De George and Wilson. This can be explained by the fact that De George and Wilson try to answer the question 'what is a good authority in a relationship between two people?' (or between groups of people, or between a written document and an individual) and not necessarily 'what is a good *argument* from authority?' In an argument from authority, the authority is cited or quoted. In that case, the presumed opinion should be accurately inferred from the actual assertion. In addition, the presumed opinion should connect adequately to the claim defended. As De George and Wilson do not put authority in such an argumentative context, they are not concerned with the accuracy of the cited opinion or the relevance of the opinion to the conclusion.

In conclusion, the argumentation-theoretical approach to authority in the realm of knowledge appears to correspond to a considerable extent to the approaches of De George and Wilson, who attempted to clarify the nature of authority from a philosophical and social epistemological perspective respectively. So, with respect to what should make a strong expert opinion, argumentation theory is not alone in its norms.

Table 2.9 Comparison between De George (1985), Wilson (1983) and argumentation theory.

<i>De George (1985)</i>	<i>Wilson (1983)</i>	<i>Argumentation theory</i>
The knowledge criterion	-	-
The inductive criterion	Occupational specialization rule Formal education rule Reputation among peers rule Reputation among other cognitive authorities rule Common consent rule Performance rule Intrinsic plausibility rule	(I.1.1) Expert criterion (I.1.1.1) Respect by other authorities criterion (I.4) External consistency criterion (I.5) Internal consistency criterion (I.6) Ability to provide evidence criterion (I.7) Recency criterion
The relevance criterion	-	(I.1) Relevant expertise criterion
The trustworthiness criterion	Personal trust rule	(I.2) Trustworthiness criterion
-	-	(I.3) Accuracy of the cited opinion criterion
-	-	(I.8) Opinion-to-conclusion relevance criterion

## 2.3 Laymen criteria

In section 2.2, I discussed argumentation-theoretical conceptions of the quality of the argument from authority. In this section, the central question is: which criteria are used by laymen to evaluate the quality of the argument from authority? A method to uncover laymen criteria had not been developed yet. Therefore, a methodological study was conducted. A more detailed report on this study can be found in Šorm, Timmers and Schellens (2007).

### 2.3.1 Methodological study

The issue of revealing laymen theories of argument quality is complex and multi-faceted. One cannot simply ask laymen to give their criteria,

because they are often not aware of them. Laymen need to be stimulated to think about the criteria they use to evaluate argument quality. Such input may come in the form of concrete arguments or situations, encouraging laymen to think about argument quality and formulate their criteria. In a methodological study, two types of stimulus material were used: a closed case and an open case. Apart from the two types of material, two different data gathering instruments were used: focus groups and individual interviews. This led to the following research question: which of the two methods, focus groups or individual interviews, yields a more detailed picture of laymen criteria?

The closed case consisted of a claim and seven supporting arguments. The arguments were manipulated in such a way that each failed to meet a different criterion that has been suggested in argumentation theory. For example, this argument violated the criterion that the cited statement should be recent: "As early as the 1950s, Robert Oppenheimer, the father of the nuclear bomb, said that nuclear energy is a promising way to reduce the greenhouse effect". The argument supported the claim "Greatly increasing the use of nuclear energy leads to a strong reduction of the greenhouse effect". The task of the respondents was to arrange the arguments in order of strength and to explain the ranking for each argument. From their explanations, it could be inferred what criteria they used for distinguishing stronger from weaker arguments from authority.

Using only this closed case may have considerable drawbacks. Since the respondents's comments were dependent on the arguments and the manipulations made by design, the respondents were likely to produce those criteria that were prompted by the flawed arguments. Applied to the example above, respondents were likely to use the criterion of recency because they were prompted by an argument violating that particular criterion. To get a realistic view on laymen criteria, a different type of case was needed to stimulate respondents to come up with their own arguments and to generate criteria spontaneously.

In the open case, respondents were asked to put themselves in a position in which it was natural to come up with a strong and a weak argument and to indicate why one argument is stronger than the other (and thus to give criteria for argument evaluation). Respondents were asked to imagine that they had applied for a job as an editor of a current affairs television program, whose task it was to invite guests for the program. They were given a fictitious claim and the task to name one person who could defend the claim properly

and one person who could not defend the claim properly. Afterwards, they had to explain why one guest was more suitable for defending the claim than the other, thus giving their criteria for the quality of the argument from authority.

We suspected that it would be a difficult task for people to come up with arguments and evaluations on their own, especially if they could not benefit from other people's ideas. A possible solution for this problem was the use of focus groups. Focus groups are group discussions facilitated by a moderator. The most important advantage of a focus group is that respondents can benefit from each other's input. This interaction stimulates ideas that would not have come up otherwise (Krueger & Casey, 2000). It is assumed by Hartman (2004) and Greenbaum (1998) that the synergistic effect generated by focus groups can be more revealing than the sum of individual interviews.

A factor that complicates focus groups is the individual characteristics of the respondents (Krueger & Casey, 2000). Dominant talkers and self-appointed experts can inhibit others in the group to put forward their opinion. The reactions of the other respondents to specific ideas or personal behavior can be influenced by one or two dominant respondents (Krueger & Casey, 2000; Greenbaum, 1998). As this is one of the biggest concerns of critics of the focus group technique, we also used individual interviews. In individual interviews, there are no negative effects of group dynamics. Respondents speak their own mind and their reactions are not influenced by others. Since a criterion mentioned by a particular respondent can be ascribed to that individual, individual interviews have the additional advantage that the agreement on a certain criterion can be determined.

A total of 48 respondents, all students at Dutch universities, participated in this study. They received an open and a closed case. The open case was always presented first, to prevent the answers in the closed case from influencing the reactions in the open case. The cases were dealt with in one of two settings: a focus group or an individual interview. 24 respondents participated in the individual interviews and 24 respondents participated in the focus groups (four focus groups, six respondents in each group).

A content analysis of the data, consisting of transcribed conversations, revealed a total of 32 different laymen criteria. The results showed that the individual interviews generated all 32 criteria, whereas focus groups generated 28 out of 32 criteria. So, the individual interviews generated four 'exclusive' criteria that were not generated by the focus groups, whereas all criteria generated by the

focus groups were also generated by the interviews. So, the individual interview appeared to be a more productive instrument than the focus group for determining laymen criteria.

Furthermore, we expected that the advantages of the focus group would be strongest in the open case, because respondents can benefit from each other's suggestions. However, in the focus groups, 20 out of 32 criteria were mentioned in the open case, whereas in the individual interviews, 25 out of 32 criteria were mentioned in the open case. Hence, the expected advantage of focus groups did not appear in the results of our study.

We also found that the open case generated 28 out of 32 criteria, whereas the closed cases generated 27 out of 32 criteria. Thus, in the open cases, five 'exclusive' criteria were mentioned that were not mentioned in the closed cases, whereas the closed cases yielded four 'exclusive' criteria that were not yielded by the open cases. So, both types of stimulus material appeared (approximately) to be equally productive.

It was concluded from the study that all instruments and material serve the purpose of finding laymen criteria. However, the individual interviews, in which both open and closed cases are used, appeared to yield a more detailed picture of laymen criteria. Furthermore, the expected advantage of focus groups, that respondents can benefit from each other's suggestions in the open case, did not appear in the results of the study. The implication for the research project was that to uncover laymen criteria, it was best to work with individual interviews on an open and closed case.

### **2.3.2 What criteria do laymen use?**

In the methodological study described in paragraph 2.3.1, the focus was on the argument from authority. This not only enabled us to evaluate methods, but also to answer the question which criteria laymen use for evaluating the argument from authority. This question has been answered in Šorm et al. (2007) to some extent, but as the main purpose in that article was to evaluate methods, less attention went to the criteria generated to evaluate the argument from authority.

### 2.3.2.1 Method

#### Material

*Open case (see appendix 2A-B).* In the open case, no arguments were given to the respondents. The respondents got the following assignment: “You have applied for a job as an editor for NOVA (a well-known Dutch current affairs program). You are now in your job interview. You have to demonstrate to the HR manager that you can distinguish between appropriate and inappropriate guests. Therefore, you have to invite two guests that both agree with a claim: one guest who can defend the claim properly and a guest who will not be able to defend the claim properly”. It was emphasized that it was an imaginary debate, so that also people who had died or people who could not speak Dutch could be invited. It was also stressed that if a respondent could not come up with a name, a description of the guest would do. After identifying both authorities, the respondents were asked to explain why they had chosen a particular authority as suitable or unsuitable. From their explanations, criteria could be derived.

The claim to be defended was either (1) “Television commercials aimed at children are bad” (In Dutch: “Televisiereclame gericht op kinderen is slecht”) or (2) “Television commercials aimed at children have detrimental effects” (In Dutch: “Televisiereclame gericht op kinderen heeft schadelijke effecten”). The first claim was considered to be a desirability claim and the latter a probability claim<sup>42</sup>. A distinction between these two types of claims was made, because the study was carried out in cooperation with Timmers (to appear), who focused on argument types supporting a desirability claim. It could have been the case that the laymen criteria would differ for the different types of claims. It appeared, however, that it did not make a substantial difference for the laymen criteria whether a desirability claim or a probability claim was defended (but see note 42).

*Closed case (see appendix 2C-D).* The closed case consisted of a claim and seven supporting arguments. The claim presented was either (1) “Nuclear energy is the best way to generate energy” (In Dutch: “Kernenergie is de beste manier om energie op te wekken” or (2) “Greatly increasing the use of nuclear energy leads to a strong

---

<sup>42</sup> It is debatable whether the probability-desirability-distinction was adequately operationalized. I will return to this point in the discussion.

reduction of the greenhouse effect” (In Dutch: “Het veel meer gebruikmaken van kernenergie leidt tot een sterke terugdringing van het broeikaseffect”). For the same reason as indicated for the open case, we chose a desirability claim and a probability claim.

The arguments were the presumed opinions of authorities. They were manipulated on the basis of criteria formulated in argumentation theory. In selecting the argumentation-theoretical criteria, each general class of specific criteria was represented (see also section 2.2), except for the categories labelled ‘ability to provide evidence criterion’ and ‘opinion-to-conclusion relevance criterion’. Table 2.10 shows the evaluation criteria violated in the closed-case arguments.

Table 2.10 Evaluation criteria violated in the closed-case arguments (translated from Dutch).

		<i>Desirability claim</i>	<i>Probability claim</i>
		Nuclear energy is the best way to generate energy.	Greatly increasing the use of nuclear energy leads to a strong reduction of the greenhouse effect.
<i>Criteria violated</i>		<i>Arguments</i>	<i>Arguments</i>
Violates: (1) expert criterion <sup>43</sup> (2) external consistency criterion (compare with C) <sup>44</sup>	A	Some physicists of the VU University of Amsterdam think that nuclear energy is the best way to generate energy.	Some physicists of the VU University of Amsterdam think that with nuclear energy, the greenhouse effect can be greatly reduced.
Violates: (1) recency criterion (2) external consistency criterion (compare with C)	B	As early as the 1950s, Robert Oppenheimer, the father of the nuclear bomb, said that nuclear energy is the best way to generate energy.	As early as the 1950s, Robert Oppenheimer, the father of the nuclear bomb, said that nuclear energy is a promising way to reduce the greenhouse effect.
Violates:	C	The famous soccer players	The famous soccer players

<sup>43</sup> The criterion was violated by leaving the source unidentified. According to Walton (1997), asking the following question is a way to determine whether someone is credible as an expert source: “What is *E*’s name, job or official capacity, location, and employer?” (p. 223). So, the source was made less credible as an expert source by not mentioning any names.

<sup>44</sup> It was intended to make arguments A-B and D-G weak with respect to external consistency, in comparison with argument C. After all, argument C is the only argument in the list in which three different sources are cited that have the same opinion.



(1) relevant expertise criterion		Ruud van Nistelrooij, Edgar Davids and Edwin van der Sar think that nuclear energy is the best way of generating energy.	Ruud van Nistelrooij, Edgar Davids and Edwin van der Sar think that the greenhouse effect can best be reduced by nuclear energy.
Violates: (1) internal consistency criterion (2) external consistency criterion (compare with C)	D	Ruud van Wijk, the chairman of Greenpeace Netherlands, is of the opinion that there is no better way of generating energy.	Ruud van Wijk, the chairman of Greenpeace Netherlands, is of the opinion that the greenhouse effect can best be suppressed by switching to nuclear energy.
Violates: (1) trustworthiness criterion (2) external consistency criterion (compare with C)	E	Frank Verbeet PhD, as a nuclear physicist connected to the nuclear power plant in Borssele, has been of the opinion for years that nuclear energy is the best way to generate energy.	Frank Verbeet PhD, as a nuclear physicist connected to the nuclear power plant in Borssele, has been of the opinion for years that nuclear energy is the best way to suppress the greenhouse effect.
Violates: (1) expert criterion (2) external consistency criterion (compare with C)	F	Wil Derben, a 17-year-old VMBO <sup>45</sup> pupil, also thinks that nuclear energy is the best way of generating energy.	Wil Derben, a 17-year-old VMBO pupil, also thinks that nuclear energy is the best solution to the greenhouse effect.
Violates: (1) accuracy of the cited opinion criterion <sup>46</sup> (2) external consistency criterion (compare with C)	G	Prof. G. Verhoeven PhD, professor in nuclear physics at the Delft University of Technology, thinks that, under certain circumstances, nuclear energy can be a fairly good way to generate energy.	Prof. G. Verhoeven PhD, professor in nuclear physics at the Delft University of Technology, thinks that under certain circumstances, nuclear energy can make a certain contribution to the fight against the greenhouse effect.

<sup>45</sup> VMBO (voorbereidend middelbaar beroepsonderwijs) is an educational level in the Netherlands, literally translated as preparatory middle-level vocational education.

<sup>46</sup> The criterion was violated by making the statement unclear and ambiguous. After all, if a statement is unclear and ambiguous, one might wonder if the cited opinion was correctly inferred from what the source actually asserted.

The respondents were asked to rank the arguments from 1 (the best argument to support the claim) to 7 (the worst argument to support the claim). After the respondents had ranked the arguments, the interviewer asked why a particular argument was put first, why another argument was put second and so on. Criteria could then be derived from their explanations.

## **Respondents**

The total number of respondents in this study was 48. 24 respondents participated in the individual interviews and 24 respondents participated in the focus groups. All 48 respondents were students at Dutch universities. The respondents were approached personally by the interviewers. Most respondents were female (77%). The ages of the respondents varied between 18 and 28, and the average age was 21 ( $SD = 2.40$ ). To make sure that the respondents could be qualified as laymen, we verified whether they had taken courses in argumentation theory. Those who had, were not included in the research sample.

## **Design**

Each respondent received one open case and one closed case. The open case was always presented first, to make sure that the reactions in the open case would not be affected by the answers in the closed case. A respondent saw either desirability claims or probability claims. So, if someone had seen a desirability claim in the open case, he or she would also see a desirability claim in the closed case. However, for each respondent, the claim topic in the open case (television commercials) was different from the claim topic in the closed case (nuclear energy).

Furthermore, the material was presented to a respondent in two different settings: either in an individual interview or in a focus group with five other respondents. In total, 24 individual interviews and four focus groups were conducted. Each focus group consisted of six respondents. This size corresponds to the ideal size of a focus group for non-commercial topics, which is usually between six and eight participants (Krueger & Casey, 2000). This design leads to four different combinations:

- 1) open case/ desirability claim 1; closed case/ desirability claim 2; individual interviews ( $n = 12$ )

- 2) open case/ desirability claim 1; closed case/ desirability claim 2; focus groups ( $n = 12$ ; 2 focus groups \* 6 respondents)
- 3) open case/ probability claim 1; closed case/ probability claim 2; individual interviews ( $n = 12$ )
- 4) open case/ probability claim 1; closed case/ probability claim 2; focus groups ( $n = 12$ ; 2 focus groups \* 6 respondents)

## **Procedure**

Each focus group was directed by a moderator team, which consisted of a moderator and an assistant moderator. The moderator was primarily concerned with directing the discussion, while the assistant helped the moderator to keep the conversation flowing. Each individual interview was conducted by only one interviewer. The interviewers and moderators were MA-students in Business Communication Studies at the Radboud University Nijmegen (the Netherlands).

The individual interviews and the focus groups followed a fixed pattern. First, the respondents were welcomed and told a fictitious purpose of the study ("to gain insight into the process of arguing"). Second, the moderator emphasized that there were no incorrect answers and that all respondents should feel free to share their points of view. Furthermore, the respondents were asked permission for the use of recording equipment and were told that their comments were confidential.

After the introduction, the interviewer or moderator presented the cases to the respondents. The open case was presented first, followed by the closed case. After reading a case, the respondents were given some time to reflect on it. In the open case, the respondents were asked to name one appropriate authority and one inappropriate authority. After naming the authorities, they were asked to explain their choices. In the closed case, the respondents were given a piece of paper and asked to rank the arguments and write down the result. After they had finished writing, all respondents were asked why they had ranked a certain authority first, second and so on.

In the focus groups, the moderator used some techniques to move the conversation along (Krueger & Casey, 2000). First of all, the moderator used the 'pause-and-probe- technique' to obtain additional information from the respondents. A five-second pause was used to elicit additional points of view or agreement with the position that was mentioned previously. The probe was a request for additional information, for instance: "Could you explain further?" and

"Is there anything else?". In addition, the moderator made sure that everyone had a chance to share, by calling on the shy respondents and asking dominant talkers to give others a chance to speak.

The mean duration of a focus group was approximately 50 minutes, whereas the mean duration of an individual interview was 15 to 20 minutes. All conversations were recorded on tape and transcribed afterwards.

### Data-analysis

The data consisted of transcripts, on which a content analysis was performed. The data were analyzed according to the steps of qualitative data analysis, as distinguished by Baarda, De Goede and Teunissen (1995).

The first step in the data analysis was to select relevant data and delete data that was irrelevant, such as utterances related to the task instruction (e.g. "Is everything clear so far?") or utterances irrelevant to the task (e.g. "Yeah, you're four years older than I am").

The second step was to divide the relevant text into fragments. The rules for dividing a fragment were: (1) The fragment should be about one topic (e.g. the evaluation of one particular argument), (2) the fragment should be readable and understandable independent from the context and (3) the fragment should not be too short to avoid taking it out of context. This, for example, was considered a fragment:

(Evaluated argument G:

Prof. G. Verhoeven PhD, professor in nuclear physics at the Delft University of Technology, thinks that under certain circumstances, nuclear energy can make a certain contribution to the fight against the greenhouse effect.

Supported claim:

Greatly increasing the use of nuclear energy leads to a strong reduction of the greenhouse effect.)

In English:

R: He is a professor in nuclear physics, so I think he knows about it, that he is objective and he says "under certain circumstances a certain contribution", so he is not very exaggerated in his statement and that's why I find him pretty realistic.

In Dutch:

Hij is een hoogleraar kernfysica, dus ik denk dat hij er verstand van heeft, dat hij objectief is en hij zegt "onder bepaalde omstandigheden een zekere bijdrage", dus hij is niet heel erg overdreven in zijn stelling en daardoor vind ik hem ook vrij realistisch.

The third step was to check each fragment, to see whether or not any criteria for evaluation were used. A criterion was defined as a norm applying to *all* arguments from authority, to avoid finding criteria dependent on the particular context in this study<sup>47</sup>. For example, if a respondent explained that he or she evaluated an argument positively because the authority does a lot of research, the decision was made that it could not be considered a criterion according to the definition. After all, the authority in question doing a lot of research was apparently inherent in the particular scientific context created in the study. By contrast, if someone explained that the positive evaluation was the result of the authority knowing about a certain subject (as in the exemplary fragment above), then that was considered a criterion according to the definition, as knowing about the subject could have been a requirement for *all* arguments from authority.

To illustrate this particular step, I will present the above fragment again, but now phrases are indicated in which criteria are used. *Which* criteria are used has also been indicated:

R: He is a professor in nuclear physics, so [1] I think he knows about it [1], [2] that he is objective [2] and [3] he says "under certain circumstances a certain contribution", so he is not very exaggerated in his statement [3] and [4] that's why I find him pretty realistic [4].

- [1] Does he know about it?
- [2] Is he objective?
- [3] Is he exaggerated in his statement?
- [4] Is he pretty realistic?

This shows that in this phase of the analysis, we stayed close to the literal verbalizations in formulating the criteria.

The fourth step was to reduce and to rearrange the set of criteria that was acquired in the previous step of the analysis. The questions that were registered in the previous step of the analysis, were grouped on the basis of a shared characteristic. The questions presented above ended up in different categories because they appeared to refer to different dimensions of the argument from authority.

The fifth step was to check whether the set of criteria was valid by using half of the data to come to a first set of criteria and see if that first set could be applied to the second half.

---

<sup>47</sup> It is questionable whether or not this was a good decision, as the consequence may have been that valuable data were not analyzed. I will discuss this issue in the final paragraph of this chapter (2.4).

## Interrater agreement

To assess the extent to which fragments could be reliably related to criteria, interrater agreement was determined. The first raters were Timmers (to appear) and myself. Timmers analyzed the data regarding desirability claims; I analyzed the data concerning probability claims. The second rater was a MA-student in Business Communication Studies at the Radboud University Nijmegen (the Netherlands), who had knowledge in the field of argumentation theory. First, she received a written task description. This description stated that she was to judge a number of fragments, in which phrases were marked. In addition, it explained that her task would be to label the phrases with criteria. The rater received a list of possible criteria she could choose from. Each criterion on the list was clarified by definitions and illustrating examples. In the instruction, an example was also presented:

- (1)<sup>48</sup> R4: Yes I always think of him as someone who follows other opinions.  
(2) R3: I always think of him as someone using difficult words. Not really interesting to listen to. Makes me think: "whatever".  
R1: He always breaks up his eight-word-sentence into three parts.
- (1) Code: Criterion:  
(2) Code: Criterion:

(In Dutch:

- (1) R4: Ja die vind ik altijd dat die een beetje meegaat met andermans mening.  
(2) R3: Die vind ik altijd een beetje moeilijke woorden hebben, zeg maar. Niet echt boeiend om naar te luisteren. Dan heb je al zoiets van "laat maar".  
R1: Hij breekt altijd z'n zin van acht woorden in drie stukjes.
- (1) Code: Criterium:  
(2) Code: Criterium:)

The second rater also received a list with rules that were used by the first raters, for instance: 'If the respondent uses a term that occurs literally in the list of criteria but interprets this term differently than we do, then choose the criterion that matches the respondent's explanation'.

---

<sup>48</sup> In the original instruction, the phrases were marked by different colours, not by different numbers.

Five training items were presented, to give the second rater the opportunity to familiarize herself with the instrument and to ask for clarification. Afterwards, 49 randomly selected test items were presented.

The results showed that the Kappa Measure of Agreement value was .804 with a significance of  $p < .0005$  ( $N = 79$ ). According to Peat and Barton (2005, p. 268), this value generally represents good agreement. A very good agreement was nearly reached, as values above .81 are generally regarded as showing very good agreement (Peat & Barton, 2005, p. 268).

### 2.3.2.2 Results

Next, I will discuss the laymen criteria that were used. Criteria that do not correspond to theory are in italics.

#### Relevant expertise criterion

Laymen in this study demand that the authority should be specialized in the field to which his assertion is related. This is demonstrated in fragment (1):

Fragment (1)
--------------

Evaluated argument (C):

The famous soccer players Ruud van Nistelrooij, Edgar Davids and Edwin van der Sar think that the greenhouse effect can best be reduced by nuclear energy.

Supported claim:

Greatly increasing the use of nuclear energy leads to a strong reduction of the greenhouse effect.

In English:

R: C, with the help of well-known football players an argument is given, whereas they, yes perhaps they know their facts, but they are not really specialized. So I would not refer to that so quickly. It don't think that's really persuasive.

I: In what way do you not find that really persuasive?

R: Because... They are football players, they are known for that. If they are also specialized in this field, you should say so. Also specialized in nuclear energy.

In Dutch:

R: C, met behulp van bekende voetballers wordt er een argument gegeven, terwijl zij, ja misschien hebben ze wel verstand van zaken, maar ze zijn er niet echt in gespecialiseerd. Dus ik zou daar niet zo gauw op terug verwijzen. Ik denk niet dat dat echt overtuigend is.

I: Hoezo vind je dat niet echt overtuigend?

R: Omdat... zij zijn voetballers, daar staan ze bekend om. Als zij ook gespecialiseerd zijn in dit vakgebied dan zou je dat erbij moeten zeggen. Ook gespecialiseerd in kernenergie.

Fragment (1) indicates that the argument would have been more persuasive if an additional statement had been made that the soccer players were specialists in the field of nuclear energy<sup>49</sup>. Thus, laymen question whether the expertise of an individual is related to the field that the assertion is in.

#### Expert criterion

Respondents in this study appeared to determine if the cited authority was a genuine expert. This is illustrated in fragment (2):

Fragment (2)
--------------

Evaluated argument (A):

Some physicists of the VU University of Amsterdam think that with nuclear energy, the greenhouse effect can be greatly reduced.

Supported claim:

Greatly increasing the use of nuclear energy leads to a strong reduction of the greenhouse effect.

In English:

R: Okay, I find A, C and F bad arguments, because A it says some physicists of the VU University of Amsterdam. I don't find that specific enough, it cannot be checked. Checked out. I find that if you refer to something, that you should be clear about what you are referring to.

In Dutch:

Oké, ik vind A, C en F slechte argumenten, omdat A er staat sommige natuurkundigen van de Vrije Universiteit van Amsterdam. Dat vind ik niet specifiek genoeg, dat valt niet na te trekken. Na te checken. Ik vind dat als je ergens naartoe refereert dat je ook duidelijk moet zijn waar naar toe.

The respondent in fragment (2) states that it is impossible to track 'some physicists'. For that reason it is impossible to determine whether the physicists are really experts or not (although the respondent does not state this consequence explicitly).

#### Trustworthiness criterion

In evaluating the argument from authority, respondents concentrated on the trustworthiness of the cited authority. Some refer to bias, as illustrated in fragment (3):

---

<sup>49</sup> The choice of the word 'persuasive' suggests that this particular respondent in fact ranked the arguments on the basis of expected persuasive effectiveness, rather than argument quality. In the data, more indications were found that expected persuasiveness played a role in the evaluations of the respondents. I will discuss this issue later.



Fragment (3)

Evaluated argument (G):

Prof. G. Verhoeven PhD, professor in nuclear physics at the Delft University of Technology, thinks that under certain circumstances, nuclear energy can make a certain contribution to the fight against the greenhouse effect.

Supported claim:

Greatly increasing the use of nuclear energy leads to a strong reduction of the greenhouse effect.

In English:

R: Number 1 is G.

I: Why do you find that one the best?

R: Because he is independent, he does not have any interests in saying if he is for or against it. And because he is not connected to an authority that takes advantage of using nuclear energy, so... and the more independent, the better. He is more capable of judging advantages and disadvantages that way. It is less subjective.

In Dutch:

R: Nummer 1 is G.

I: Waarom vind je die het allerbeste?

R: Omdat hij onafhankelijk is, hij heeft er geen belangen bij om te zeggen of hij voor of tegen is. En omdat hij niet verbonden is aan enige instantie die voordelen heeft bij het gebruik van kernenergie, dus... en hoe onafhankelijker, hoe beter. Zo kan hij beter de voor- en nadelen beoordelen. Het is minder subjectief.

This respondent indicates that the evaluated argument is best, because the authority can consider the issue objectively. Others referred to honesty, as the following fragment shows:

Fragment (4)

Evaluated argument (D):

Ruud van Wijk, the chairman of Greenpeace Netherlands, is of the opinion that the greenhouse effect can best be suppressed by switching to nuclear energy.

Supported claim:

Greatly increasing the use of nuclear energy leads to a strong reduction of the greenhouse effect.

In English:

I: Okay. And whom would you put on number 3?

R: Er, I think D.

I: Greenpeace, and why?

R: Yes he is sympathetic towards the environment, of course Greenpeace is only a er...yes...I am not all that happy with them and stuff, but I think that he is being honest in this case.

I: What do you mean he is being honest in this case?

R: Yes, sometimes they tend to... that they only act according to their own interests and not to you know the common good. And as I read it like that, I think he is being sincere.

In Dutch:

I: Oké. En wie zou je op nummer 3 plaatsen?

R: Eh, ik denk D.

I: Greenpeace, en waarom?

R: Ja die is vrij begaan met het milieu, natuurlijk is Greenpeace ook maar een eh... ja... ik ben er niet geweldig over te spreken of zo, maar ik denk dat hij in dit geval wel eerlijk is.

I: Hoezo is hij in dit geval eerlijk?

R: Ja, soms zijn ze nogal... dat ze alleen gaan voor wat ze zelf vinden en niet voor het grootste belang zeg maar. En volgens mij zoals ik het zo lees is hij wel oprecht.

---

### External consistency criterion

Respondents question whether what the cited authority says is in agreement with what others say. Consider fragment (5):

Fragment (5)
--------------

Evaluated argument (A):

Some physicists of the VU University of Amsterdam think that with nuclear energy, the greenhouse effect can be greatly reduced.

Supported claim:

Greatly increasing the use of nuclear energy leads to a strong reduction of the greenhouse effect.

In English:

R: Then I put A in 3<sup>rd</sup> place. I didn't put it on 1, because there it also says "physicists of the VU University", but here it says "some". That means that others disagree with it. So that's not really much good to me.

In Dutch:

R: Dan heb ik op 3 A staan, die heb ik niet op 1 gezet, want daar staan ook weer "natuurkundigen van de Vrije Universiteit", maar hier staat "sommige". Dat betekent dat anderen het er weer niet mee eens zijn. Dus daar heb ik niet zoveel aan.

---

To the respondent, the phrasing "some physicists" implies that their opinion is not shared by all physicists, which leads to a negative evaluation of the argument.

### Internal consistency criterion

Laymen appear to refer to internal consistency, as fragment (6) illustrates:

Fragment (6)
--------------

Evaluated argument (E):

Frank Verbeet PhD, as a nuclear physicist connected to the nuclear power plant in Borssele, has been of the opinion for years that nuclear energy is the best way to suppress the greenhouse effect.

---

Supported claim:

Greatly increasing the use of nuclear energy leads to a strong reduction of the greenhouse effect.

In English:

R: After that, E, also a nuclear physicist, who has had that opinion for years, I find that a reliable source as well. I would believe someone like that.

In Dutch:

R: Daarna dan, E, ook kernfysicus, die heeft al jarenlang die mening, dat vind ik ook een betrouwbare bron, dat neem ik wel aan van zo iemand.

---

In fragment (6) the cited Frank Verbeet is seen as a reliable source, because he has held the same opinion for years. In other words, the judgment that the opinion is internally consistent leads to a positive evaluation of the argument.

### Flexibility criterion

The decision that an argument from authority has a certain property may result in a negative or a positive conclusion about the argument's quality. Consider the following fragment:

Fragment (7)
--------------

Evaluated argument (E):

Frank Verbeet PhD, as a nuclear physicist connected to the nuclear power plant in Borssele, has been of the opinion for years that nuclear energy is the best way to suppress the greenhouse effect.

Supported claim:

Greatly increasing the use of nuclear energy leads to a strong reduction of the greenhouse effect.

In English:

I: Okay, in the middle part of the ranking if I may say so was that 17-year old VMBO-pupil, Robert Oppenheimer, and Dr. Frank Verbeet. Hey, that's interesting since he is also a nuclear physicist. Just like Dr. G. Verhoeven. Then why not Dr. Verbeet?

R: I get that, but the difference between the two is, he is connected to the nuclear power plant in Borssele. In other words: he is defending his own job. And secondly, he has been claiming for years that nuclear energy is the best way to suppress the greenhouse effect. It's a bit too obvious for me. I just don't think it's reliable. It could have been possible that if that man had had another job, and he wasn't so obvious, that I would have thought he was reliable. But this just doesn't seem to be an unbiased opinion, so...

In Dutch:

I: Oké, het middengedeelte als ik het zo mag zeggen van de rangorde, dat was die 17-jarige VMBO-scholier, Robert Oppenheimer en dr. Frank Verbeet. Hé, dat is opmerkelijk want dat is ook een kernfysicus. Net als dr. G. Verhoeven. Waarom dr. Verbeet dan niet?

R: Ik snap je wel, maar het verschil tussen die twee is, hij is verbonden aan de kerncentrale Borssele. Met andere woorden: hij staat zijn eigen baan te verdedigen. En ten tweede is hij al jaren van mening dat kernenergie de beste methode is om het broeikas effect te bestrijden. Dat ligt er allemaal dik bovenop. Dus dat vind ik gewoon niet betrouwbaar. Het kan best zijn dat als die man een andere baan had gehad en iets minder het dik er boven op had gelegd dat ik hem dan wel betrouwbaar had gevonden. Maar dit vind ik gewoon niet onafhankelijk overkomen. Dus.

---

Similar to the previous fragment, this respondent mentions that the source Frank Verbeet has had the same opinion for years (see underlined part). However, instead of regarding that as a positive characteristic, the respondent interprets that as a sign for being an unreliable source. It is not internal consistency that appears to be the criterion here, but rather flexibility, or the source's capability of changing opinions from time to time.

#### Ability to provide evidence criterion

In evaluating the argument from authority, the respondents refer to the evidence to explain or justify the opinion. For instance:

Fragment (8)
--------------

Evaluated argument (F):

Wil Derben, a 17-year-old VMBO pupil, also thinks that nuclear energy is the best solution to the greenhouse effect.

Supported claim:

Greatly increasing the use of nuclear energy leads to a strong reduction of the greenhouse effect.

In English:

R: I don't find a 17-year old VMBO-pupil, in argument F, very convincing either. Look, if he had added a fantastic argument, but it says nuclear energy is the best solution to the greenhouse effect. Well, that's already in the claim.

In Dutch:

R: Ik vind een 17-jarige VMBO scholier, bij argument F, ook niet echt overtuigend. Kijk als hij er nou een geweldig argument bij had gezet, maar er staat de kernenergie is de beste oplossing voor het broeikas effect. Tsjja, dat staat ook in de stelling.

---

The respondent indicates that this particular argument from authority would have been more persuasive, if the pupil Wil Derben had

backed up his opinion with arguments. That is quite remarkable, as in each argument in the closed case, the opinion was not backed up with arguments. It is possible that laymen particularly want individuals who have relatively low credibility as genuine experts, to back up their assertions.

#### Recency criterion

It appears from this study that laymen demand that the statement of the authority is not outdated. Consider fragment (9):

Fragment (9)
--------------

Evaluated argument (B):

As early as the 1950s, Robert Oppenheimer, the father of the nuclear bomb, said that nuclear energy is a promising way to reduce the greenhouse effect.

Supported claim:

Greatly increasing the use of nuclear energy leads to a strong reduction of the greenhouse effect.

In English:

R: I ranked A above B. It's nice 'n all that he made that nuclear bomb, but... we are not exactly in the early 1950's anymore but have come a lot further along. And I'm not sure if this guy has made the effort to stay up to date on the subject.

In Dutch:

R: Ik had A boven B. Leuk dat hij die atoombom heeft gemaakt, maar intussen... we zijn nu niet in begin jaren 50, maar intussen zijn we een stuk verder. En ik weet niet of je dan nog moeite neemt om de boel bij te spijkeren.

The respondent notes that the assertion is outdated and therefore, he questions if Robert Oppenheimer himself is an up-to-date expert.

#### Ahead-of-his-time expert criterion

An assertion made years ago can also be taken as a clue for the argument being strong. Consider fragment (10):

Fragment (10)
---------------

Evaluated argument (B):

As early as the 1950s, Robert Oppenheimer, the father of the nuclear bomb, said that nuclear energy is a promising way to reduce the greenhouse effect.

Supported claim:

Greatly increasing the use of nuclear energy leads to a strong reduction of the greenhouse effect.

In English:

I: Okay, you put B Robert Oppenheimer in third place. Okay, can you describe once again why you put Oppenheimer in third place? That is also a fairly strong

argument if you can see that in the ranking. Why did you put him in third?

R: He also seems a man who knows his stuff, a nuclear bomb and he works with nuclear energy and its waste. Yes and also a little historical perspective that he already understood that back then. If that can still be maintained today then it could be a strong argument or fairly strong.

In Dutch:

I: Oké, en op nummer drie heb je dan B Robert Oppenheimer. Oké, kun je dan ook weer aangeven waarom je Oppenheimer op nummer drie weer hebt neergezet? Dat is dan ook nog wel een redelijk sterk argument als je dat in de rangorde kunt zien. Waarom heb je hem op nummer drie gezet?

R: Het lijkt me ook wel weer een man die er wat van af weet, een atoombom en die heeft met kernenergie te maken en afval ervan. Ja en ook een stukje historisch perspectief dat hij dat toen al doorzag eigenlijk. Als dat nu nog overeind kan staan dan zou dat wel een sterk argument kunnen zijn of redelijk sterk.

---

The respondent in fragment (10) suggests that it can be a good thing that a statement is made some time ago – under the condition that the statement still holds - because it means that for the source, the issue was obvious already years ago. In other words: the source was ahead of its time.

#### Hedged statement criterion

Laymen in this study pay attention to the certainty expressed in the assertion. Some see uncertainty as a reason to give the argument a positive evaluation:

Fragment (11)
---------------

Evaluated argument (G):

Prof. G. Verhoeven PhD, professor in nuclear physics at the Delft University of Technology, thinks that under certain circumstances, nuclear energy can make a certain contribution to the fight against the greenhouse effect.

Supported claim:

Greatly increasing the use of nuclear energy leads to a strong reduction of the greenhouse effect.

In English:

I: Which argument would you use to defend the claim?

R: I think the letter G, of Prof. Dr. G. Verhoeven.

I: And why exactly?

R: Well, when I read him, he seemed the most reliable to me, the most convincing also because, let's see, when you look at A, there it is presented really simply, they think it can be reduced and that's that. Then there, it's somewhat more specific, that under certain circumstances, and then with the help of nuclear... that's formulated a little more conditionally. That just sounds a little more reliable than just saying it can be done with...

I: Okay, so do you mean the fact it is somewhat more specific, makes the argument more reliable?

R: Yes because it seems more precise and yeah... yeah.

In Dutch:

I: Welke argument zou jij gebruiken om de stelling te verdedigen?

R: Ik denk letter G, van Prof. Dr. G. Verhoeven.

I: En waarom precies?

R: Ja, toen ik hem zo las, kwam die mij het meest betrouwbaar over, het meest overtuigend ook omdat, even kijken, als je bij A kijkt, daar wordt het heel simpel gebracht, die denken dat het terug kan worden gedrongen en klaar. Dan daar is het toch iets specifieker dat onder bepaalde omstandigheden, en dan met behulp van kern... dat is iets voorwaardelijker gebracht. Dat klinkt gewoon wat betrouwbaarder dan dat je gewoon zegt dat lukt wel met...

I: Oké, dus bedoel je dat dat specifieke, dat maakt dat argument betrouwbaar?

R: Ja omdat het dus nauwkeuriger lijkt en ja... ja.

---

### Certain statement criterion

Others would rather like to see a certain statement:

Fragment (12)
---------------

Evaluated argument (E):

Frank Verbeet PhD, as a nuclear physicist connected to the nuclear power plant in Borssele, has been of the opinion for years that nuclear energy is the best way to suppress the greenhouse effect.

Supported claim:

Greatly increasing the use of nuclear energy leads to a strong reduction of the greenhouse effect.

In English:

I: Okay, you told me a couple of things, but what is the main reason you chose E and not G or D, like the others did? What do you think are the most important differences between them?

R: The combination of knowledge of theory and practice as a matter of fact. And the clarity with which the argument is formulated, so that there is just no doubt.

In Dutch:

I: Oké, je hebt het al een beetje verteld, maar wat is voor jou de grootste reden om voor E te kiezen en niet voor G of D die de anderen hebben gekozen? Wat zijn daar voor jou de belangrijkste verschillen tussen?

R: Dus inderdaad de combinatie van kennis van theorie en praktijk. En de duidelijkheid waarmee het argument wordt gebracht, gewoon dat er geen twijfel in zit.

---

So, the fragments (11) and (12) show that certainty of the statement may lead to opposite conclusions: the argument from authority is good or bad.

Table 2.11 shows the criteria that laymen used to evaluate the quality of the argument from authority. Criteria that do not correspond to theory are in italics.

Table 2.11 Laymen criteria for the argument from authority (per type of claim, data gathering instrument, type of stimulus material) and agreement among respondents (the number of individual interviews in which the criterion was mentioned divided by the total number of individual interviews) (laymen criteria that do not correspond to theory are in italics).

Criterion	Probability claim				Desirability claim				Agreement among respondents in individual interviews
	Individual interviews		Focus groups		Individual interviews		Focus groups		
	Open case	Closed case	Open case	Closed case	Open case	Closed case	Open case	Closed case	
Relevant expertise	+	+	+	+	+	+	+	+	15 of 24 (63%)
Expert	+	+	+	+	+	+	+	+	24 of 24 (100%)
Trustworthiness	+	+	+	+	+	+	+	+	22 of 24 (92%)
External consistency	+	+	-	-	-	+	-	+	2 of 24 (8%)
Internal consistency	+	+	-	-	-	+	+	+	5 of 24 (21%)
<i>Flexibility</i>	-	+	-	-	-	-	+	-	2 of 24 (8%)
Ability to provide evidence	+	+	+	+	+	+	+	+	14 of 24 (58%)
Recency	-	+	-	+	-	+	-	+	13 of 24 (54%)
<i>Ahead-of-his-time expert criterion</i>	-	-	-	-	-	+	-	-	1 of 24 (4%)
<i>Hedged statement</i>	+	+	-	+	-	+	-	+	16 of 24 (67%)
<i>Certain statement</i>	-	+	-	+	+	+	-	+	3 of 24 (13%)



## 2.4 Conclusion and discussion

### Conclusion

The central question of this chapter is: to what extent do lay persons' criteria for the evaluation of the argument from authority correspond to the criteria formulated in argumentation theory? It appears that the evaluation standards correspond to a considerable extent. However, we should also take into consideration the extent to which laymen in this study agree on criteria that are also mentioned in theory. It appears that more than half of the respondents who participated in the individual interviews mentioned the relevant expertise criterion, the expert criterion, the trustworthiness criterion, the ability to provide evidence criterion and the recency criterion. For the external consistency criterion and the internal consistency criterion, the agreement among laymen appears to be considerably lower. In sum:

Corresponding criteria with relatively high agreement among respondents:

- The relevant expertise criterion: the source's expertise should be relevant to the opinion.
- The expert criterion: the source should be a genuine expert.
- The trustworthiness criterion: the source should be trustworthy.
- The ability to provide evidence criterion: the source should be able to provide evidence.
- The recency criterion: the source's statement should be recent.

Corresponding criteria with relatively low agreement among respondents:

- The external consistency criterion: the source's statement should be externally consistent.
- The internal consistency criterion: the source's statement should be internally consistent.

Most criteria that are distinguished in argumentation theory are used by laymen. There are some differences, however. Laymen in this study do not question the accuracy of the cited opinion, which means that they do not question whether what the expert actually asserted had been correctly understood, interpreted, quoted, etc. Although the intention was to violate this particular criterion in the closed case by making the statement unclear and ambiguous, the respondent had no reason to believe that the opinion was incorrectly cited. Instead, this manipulation stimulated respondents to focus on the certainty with which the presumed opinion was expressed. If the opinion is posed with certainty, then that may cause the argument from authority to be evaluated as either strong or weak. This,

however, was not suggested in argumentation theory for the argument from authority.

Furthermore, respondents did not use the opinion-to-conclusion relevance criterion, nor did they focus on the strength of the conclusion. This was to be expected, as the material was not manipulated on the basis of these aspects.

In addition, laymen use the criterion that the source should be flexible (in the sense of being able to change opinion) and according to one layman, using an expert who is ahead of his time should make the argument strong. This was not mentioned in argumentation theory.

If we take into consideration the agreement among participants in the individual interviews, it appears that the hedged statement criterion is used by a majority of participants, whereas the flexibility criterion, the certain statement criterion and the ahead-of-his-time expert criterion is used by a minority of participants. In sum, the non-corresponding criteria are:

Non-corresponding criteria that are not mentioned by respondents:

- The accuracy of the cited opinion criterion: the source's opinion should be accurately cited (theory only, as expected, because of material).
- The opinion-to-conclusion relevance criterion: the source's opinion should be relevant to the conclusion (theory only, as expected, because of material).
- The strength of conclusion criterion: the conclusion should not be too strong (theory only, as expected, because of material).

Non-corresponding criterion with relatively high agreement among respondents:

- The hedged statement criterion: the source's statement should be hedged (laymen only, unexpected).

Non-corresponding criteria with relatively low agreement among respondents:

- The flexibility criterion: the source should be flexible (laymen only, unexpected).
- The certain statement criterion: the source's statement should be certain (laymen only, unexpected).
- The ahead-of-his-time expert criterion: the source should be ahead of his time (laymen only, unexpected).

## Discussion

With respect to criteria for the argument from authority, argumentation theory and laymen are generally on the same line, although some

laymen criteria appear to be more agreed upon in this study than other criteria. The relevant expertise criterion, the expert criterion, the trustworthiness criterion, the ability to provide evidence criterion and the recency criterion were mentioned in the greater number of individual interviews; the external consistency criterion and the internal consistency criterion were mentioned in the smaller number of individual interviews.

Some differences between argumentation theory and laymen are revealed in this study, but these should be interpreted carefully. After all, a part of the differences can be explained by the stimulus material. It is possible that laymen had insufficient reason to believe that the cited opinion was not accurate, because of the setup of the material. If we would have added direct quotes from the expert, or transcripts of dialogues, to the opinion, it might have been possible to provoke such a criterion. The same goes for the strength of the conclusion criterion and the opinion-to-conclusion relevance criterion that were not mentioned by laymen. Had we created conclusions varying in strength and relevance, these criteria might have been provoked.

Some criteria that were only used by laymen, were only shared by a few laymen: the flexibility criterion, the ahead-of-his-time expert criterion and the certain statement criterion. This indicates that we should not place too much value on these differences. One laymen criterion that differs from theory should be given more significance: the hedged statement criterion, which was shared by 16 of 24 respondents. This indicates that quite a number of laymen in this study want the source in the argument from authority to be cautious in making unqualified statements, a requirement that is not mentioned in theory. In general, however, theory and laymen appear to have the same vision on what should make an argument from authority strong.

How can one explain that quite a number of laymen focus on the certainty with which the presumed opinion is expressed, which is not suggested in argumentation theory? This was often stimulated by an argument in the closed case that aimed to violate the accuracy of the cited opinion criterion: e.g. "Prof. G. Verhoeven PhD, professor in nuclear physics at the Delft University of Technology, thinks that under certain circumstances, nuclear energy can make a certain contribution to the fight against the greenhouse effect". Respondents generally did not perceive this as an unclear statement in the sense of 'vaguely stated' or 'unclear to the mind', but as a hedged statement. This property of being unsure resulted in either a positive

or a negative evaluation. That is also understandable: if the statement is hedged, the source gives the impression not to overlook necessary qualifications or exceptions. On the other hand, the absence of hedges may give the impression of the expert being certain about the issue.

The data in this study gave insight into some other interesting phenomena. First of all, the phenomenon that laymen may see that the argument from authority has a certain characteristic, e.g. being certain or having the same opinion for years, but then reach different conclusions about the quality of the argument. Second: laymen may use judgments in certain dimensions in coming directly to a decision on argument strength, but judgments may also be used indirectly to form judgments in other dimensions, before a final decision on argument strength is made. Some, for instance, relate having the same opinion for years directly to argument quality, whereas others relate it to trustworthiness or being an expert before they make a final assessment of argument quality. Thus, the results indicate that complex considerations may underlie a decision about argument strength.

If the laymen criteria are compared with the criteria by Wilson (1983) and De George (1985) discussed earlier in this chapter, there is considerable overlap. One difference is that laymen in this study do not use De George's (1985) knowledge criterion, expressing that there should be genuine knowledge to be had. This can be explained by the fact that the material was not manipulated on the basis of that aspect.

This study has some limitations. The data show that some respondents might have focused on expected persuasiveness in evaluating the arguments. This can be explained by the material and the instruction. In the open case, the respondents were requested to come up with two television guests to defend a certain claim. As a consequence, some might have been tempted to choose authorities differing in rhetorical capabilities. The instruction in the closed case might also have provoked a ranking on the basis of persuasive effectiveness: "To support this claim, you can choose from the arguments below. Which of these arguments would you prefer to use to defend this claim? Which arguments, on the contrary, do you consider to be very weak defending the claim? Would you be able to rank these arguments in order of 1 to 7, with 1 as the best argument and 7 as the worst?" (translated from Dutch). It is not impossible that respondents were stimulated by the phrasing 'best/worst' to distinguish between arguments on the basis of their persuasiveness

and therefore, used different criteria than they would have used if they distinguished the arguments on the basis of reasonableness.

Second, the intention was to make a distinction in the material between desirability and probability claims, to check whether there would be any differences in results between these claim types. It is questionable if the probability-desirability distinction was operationalized in a valid way. First of all, the claims that were supposed to be desirability claims would probably not occur in a pragmatic argumentation scheme as a premise, because obviously acts are evaluated instead of effects: "Television commercials aimed at children are bad" and "Nuclear energy is the best way to generate energy". Desirability claims like "Strong reduction of the greenhouse effect is a good idea" and "Children buying food high in fat, salt and sugar is bad" would have been more adequate as desirability claims, because in these claims the possible effects of proposed acts are clearly evaluated. Apart from this, the label 'probability claim' for the claim 'Television commercials aimed at children have detrimental effects' is dubious. After all, it does not express that a certain act (probably) leads to a certain consequence. It only says that the effects of the proposed act - whatever these may be - are detrimental. Therefore, the chosen claim might be considered inadequate in representing a probability claim.

Third, in the data-analysis, the decision was made to take criteria into consideration that applied to all authorities and to disregard criteria that were dependent on the context created in the material. It is questionable whether or not this was a good decision. After all, by ignoring context-dependent criteria we missed out on questions that might have served laymen as heuristics. For instance, the fact of doing a lot of research may indicate that someone knows about a certain subject, or that someone is a genuine specialist in the field. Moreover, if we had taken these criteria into consideration, we would have gained interesting insights into the evaluation of appeals to epistemic/cognitive authorities in the scientific or academic context.

All in all, we can conclude that laymen criteria for evaluating the quality of the argument from authority correspond to a considerable extent to the criteria on the basis of argumentation theory. There appears to be relatively high agreement between laymen on the relevant expertise criterion, the expert criterion, the trustworthiness criterion, the ability to provide evidence criterion, the recency criterion and the hedged statement criterion. Therefore, it can be concluded that these criteria are usually part of laymen's argumentative competence.

### 3. The quality of the argument from cause to effect

This chapter deals with the quality of the argument from cause to effect. First, in section 3.1, I will discuss causality and the conditions that should be met in order to be able to speak of a causal relationship. I will also indicate which variant of the argument from cause to effect is relevant to further investigation. Subsequently, I will address the questions of what criteria are used in argument theory to evaluate the quality of arguments from cause to effect (3.2), what criteria are used by laymen to evaluate the quality of these arguments (3.3) and finally, I will make a comparison between these sets of criteria (3.4).

#### 3.1 Introduction

##### 3.1.1 Causality conditions

The causality concept is crucial to the quality of the argument from cause to effect. For a causal relationship to be plausibly present, certain causality conditions should be satisfied. The work of Hulswit (2002) is useful<sup>50</sup> in elucidating the nature of causality and the criteria for a causal relationship. Hulswit provides an overview of contemporary philosophical accounts of causality. I will discuss four of the five approaches he deals with and leave out the so-called *singularist approach*<sup>51</sup>, since Hulswit (2002) considers it “a minority view” (p. 47).

The first approach to causality is the *necessary and/or sufficient conditions approach*, associated with the philosophy of David Hume (1739-1740, 1978) and John Stuart Mill (1874). Logically, a necessary condition for the occurrence of an effect is a circumstance without which the effect cannot occur. A sufficient condition for the occurrence of an effect is a circumstance in whose presence the effect must occur (Copi, 1982). Hulswit (2002, p. 52) calls John Mackie’s (1974) analysis in terms of *INUS conditions*<sup>52</sup> the

---

<sup>50</sup> For other reviews of approaches to causality, see White (1990) and Brady and Seawright (2004).

<sup>51</sup> This approach “is characterized by the idea that causal relations are irreducibly individual, rather than instantiations of universal relations. Thus, one may describe the cause of an event B as the one single event (A) that occurs closest to and immediately preceding event B” (Hulswit, 2002, p. xvi).

<sup>52</sup> Called after the initial letters of key-words in Mackie’s definition of cause (Insufficient but non-redundant part of unnecessary but sufficient condition).

“more sophisticated version” of an analysis in terms of necessary and sufficient conditions:

“an event A is the cause of an event B if A is a necessary part of a complex condition C, which, though sufficient, is not necessary for the effect (B)” (Mackie, as paraphrased by Hulswit, 2002, p. 73).

I can illustrate this definition by saying that good teamwork is the cause of successful sport performance, if it is the decisive (therefore: necessary) factor in a set of circumstances (including, for instance, nice weather conditions, specialized diets and good mental skills). This set of circumstances is sufficient for the occurrence of high sport performance, but not necessary, since the high sport performance could have been caused by another set of circumstances.

The second approach to causality that Hulswit (2002) describes is *the counterfactual approach*, associated with the work of David Lewis (1973, 1993). According to this approach, a cause makes a difference: if X had not happened, Y would (other things being equal) not have happened. For instance, saying that the lightning is the cause of the fire, is saying that if the lightning had not struck, the fire would never have existed. Many contemporary philosophers define causality in terms of relations of counterfactual dependence (Hulswit, 2002). Counterfactual dependence means that the cause is rendered counterfactually necessary for the effect and moreover, that A is linked to B through a causal chain of counterfactually dependent events. As Hulswit (2002) puts it: “A is the cause of B if and only if there is a chain of counterfactually dependent events linking A and B” (p. 55).

The third approach to causality that Hulswit (2002) deals with is *the instrumental or agency approach*, defended by Collingwood (1938, 1991) and Von Wright (1971). From this perspective, a cause is regarded as “an event or state that we can produce or prevent at will, or otherwise manipulate, in order to produce or prevent a certain other event as an effect” (Hulswit, 2002, p. 56). In this respect, the cause-effect relationship is related to our ability to control the world as agents and causes are seen as means to an end. For instance, given this account, the boiling of water is a cause of hot water, since boiling water is an effective means by which we, free agents, can bring about the occurrence of hot water.

The next approach discussed by Hulswit (2002), is the *probabilistic approach*, represented by Patrick Suppes (1970). According to this approach, “an event A may be said to be a cause of an event B, if, given the occurrence of A, the probability of the

occurrence of B is higher than the probability of the occurrence of B would have been if A had not occurred” (p. 73). For example, burning oil may be considered a cause of global warming if (and only if) it raises the probability of global warming.

In sum, the contemporary philosophical accounts of causality can be reduced to (I) a necessary and/or sufficient conditions approach, (II) a counterfactual approach, (III) an instrumental or agency approach and (IV) a probabilistic approach.

### 3.1.2 The argument from cause to effect

Argumentation theorists appear to agree that the argument from cause to effect concerns (a) causality and (b) prediction. Walton (1996), for instance, claims that the argument from cause to effect “takes the form of a prediction or warning that one type of event tends to cause another” (p. 73). Another illustration is the definition by Hastings (1962). According to him, reasoning from cause to effect is asserting “that because certain events exist, then certain other events can be expected to exist either simultaneously or subsequent in time as a result of the first events” (p. 65). As appears from various definitions, the argument from cause to effect is undisputedly regarded as making the prediction of a certain effect in the claim acceptable by presenting a cause in the argument that is supposed to lead to the effect.

Two forms of the argument from cause to effect are distinguished in argumentation theory. The first form is a prediction based on existing conditions. This form is schematically represented by Schellens (1985, p. 100, translated from Dutch) as:

A (generally) leads to B  
Ai  
Therefore: (probably) Bi

In this argumentation scheme, the conclusion that something will occur is based on a causal generalization that kinds of causes A lead to kinds of effects B and on the premise that a certain cause Ai is the case. Ai is a member of the category of causes A presented in the causal generalization.

The second form is a prediction based on hypothetical conditions. This type of reasoning expresses the possible consequences that may arise as a result of implementing a certain policy (Hastings, 1962; Schellens, 1985). Schellens (1985) presents two formal schemes corresponding to the hypothetical variant. In the



following scheme, an effect is predicted on the basis of a general causal relationship (p. 92, translated from Dutch):

A generally leads to B  
Therefore: if Ai, then probably Bi

In addition, a scheme is presented in which an effect is predicted on the basis of two or more conditionals (p. 92, translated from Dutch):

If Ai then (probably) Bi  
If Bi then (probably) Ci  
Therefore: if Ai then (probably) Ci

Critical questions have been introduced that should be applied to the argument from cause to effect. Sometimes, these questions are formulated exclusively for this type of argument (e.g. Schellens & Verhoeven, 1994; Walton, 1996) and sometimes the same questions are intended for effect-to-cause reasoning as well (e.g. Warnick & Inch, 1989). Furthermore, in the sets of questions, no distinction has been made between questions applying to the 'existing condition' variant and the 'hypothetical' variant. The only exception to this rule appears to be Schellens (1985, p. 100).

What forms of the argument from cause to effect are relevant within the framework of pragmatic argumentation? In chapter 1, I explained that I would focus on the evaluation of argument types supporting one of the premises in the pragmatic argumentation scheme: the premise expressing that action A (probably) leads to B. The action A in this premise is not an existing condition, but a conditional, or hypothetical one. The hypothetical variant of the argument from cause to effect is particularly relevant to pragmatic argumentation, since pragmatic argumentation is about actions that have not been executed yet (Schellens, 1985). Therefore, this study will focus on the hypothetical variant of the argument from cause to effect.

More specifically, the focus will be on the two forms that Schellens (1985) distinguishes: the form in which the conclusion is inferred from a general causal relationship and the form in which the conclusion is inferred from two or more conditional predictions. Table 3.1 shows how these forms relate to the pragmatic argumentation scheme.

Table 3.1 Forms of the argument from cause to effect relevant to pragmatic context.

<i>1. Argumentation based on a general causal relation</i>	
<i>Formal scheme</i>	<i>Example</i>
A generally leads to B Therefore: if A, then probably B  (B is desirable) (Therefore: A is desirable)	The introduction of innovative products generally leads to an easier way of living. Therefore: if the public transport chip card is introduced, then using public transport will be easier. (Easier use of public transport is desirable). (Therefore: the introduction of the public transport chipcard is desirable).
<i>2. Argumentation based on two or more conditional predictions</i>	
<i>Formal scheme</i>	<i>Example</i>
If A then (probably) B If B then (probably) C Therefore: if A then (probably) C  (C is desirable) (Therefore: A is desirable)	If the chip card is introduced, then you will need only one card for public transport. If you need only one card, then traveling will be more convenient. Therefore: if the public transport chip card is introduced, then using public transport will be more convenient. (More convenient public transport is desirable). (Therefore: the introduction of the public transport chipcard is desirable).

In section 3.2, I will present a study on the argumentation-theoretical conceptions of strong arguments from cause to effect and discuss how these correspond to the views on causality as discussed in section 3.1.1.

### **3.2 Argumentation-theoretical criteria**

In this section, the key question is: what evaluation questions have been suggested in argumentation theory for the evaluation of the argument from cause to effect? To address this question, a literature study was carried out.

#### **3.2.1 Method**

The methodological details have been discussed in chapter 2 on the argument from authority. Therefore, I refer to section 2.2.1 for a

description and justification of the selection of literature. With two exceptions, the same literature was employed in this study. First, I did not use Walton's work on the appeal from expert opinion (1997), but Walton's *Argumentation Schemes for Presumptive Reasoning* (1996) instead. The former work is exclusively about the appeal from expert opinion, whereas the latter contains different argumentation schemes and critical questions, among which questions for the argument from cause to effect.

Second, I used Freeley and Steinberg (2000). I did not use this publication in chapter 2, because they have not formulated critical questions specifically for the argument from authority. In this chapter, however, I will consider them, because they suggested questions specifically for causal reasoning.

### 3.2.2 Results

This section shows the results of the inventory and classification of evaluation questions for the argument from cause to effect and also for causal reasoning in general. In my opinion, these questions can be divided into three categories. The first category contains the questions that are specifically useful in evaluating the quality of the argument from cause to effect. The second category consists of evaluation questions that can be applied in the evaluation of the argument from cause to effect, but also in the evaluation of other argument types. In the third category, I placed criteria that can be considered specific to argument schemes other than the argument from cause to effect. It should be explicitly noted, however, that the authors who suggested these criteria did so for *causal reasoning in general, including cause-to-effect and effect-to-cause*, and not specifically for the argument from cause to effect (Freeley & Steinberg, 2000; Meany & Shuster, 2002; Reinard, 1991; Warnick & Inch, 1989). Hence, it is certainly not the case that these authors erroneously formulated criteria for the argument from cause to effect that are in fact irrelevant. They formulated criteria that may be relevant to causal reasoning in general, but that can be regarded as irrelevant to the argument from cause to effect specifically.

Within these three more general categories, subcategories can be distinguished. Next, I will discuss each of these subcategories.

## (I) Specific to the argument from cause to effect

### (I.1) Cause sufficiency criterion

Schellens (1985, pp. 96-97; 1994, p. 81) commented that it is only relevant to know whether a cause is a necessary condition, if one reasons from effect to cause. To reason from cause to effect, it should be enough to know whether a cause is a sufficient condition. Therefore, as he considered necessity as irrelevant to the argument from cause to effect, he only required that the cause be sufficient. Hence his question: "Is cause A in general sufficient to consider effect B probable?" (1994, p. 82) (see Table 3.2).

In pragma-dialectics, two questions are formulated that in my opinion both refer to sufficiency (Garssen, 1997). The first is: "Does that what is proposed as a cause (Z) indeed lead to the effect mentioned (Y)?" (p. 20, translated from Dutch). Furthermore, the pragma-dialecticians suggest that we should consider whether anything else, other than the cause under consideration, is missing for the effect to occur (Garssen, 1997). If one comes to the conclusion that other necessary causes should be present for the effect to occur, then the question of whether the cause is a sufficient condition, can be answered with a 'no'. Therefore, both questions belong to (I.1).

Warnick and Inch (1989, p. 109) and Freeley and Steinberg (2000, p. 162) demand that the cause is "capable of producing the effect"; Reinard (1991) wants it to "carry enough force to produce the effect" (p. 199). It could be argued that this is a way of saying that a cause should be sufficient for the effect, though in terms of capability, or the quality of being able to produce the effect.

I created two subcategories for tests of cause sufficiency. Firstly, in argumentation theory, critical questions are formulated regarding contributing and/or counteracting factors that influence the chance of the occurrence of the effect (I.1.1). Contributing factors raise the probability of the effect; counteracting factors lower the probability of an effect. Thus, to determine if a cause is enough to bring about the effect, people should look for causal factors that modify the effect's occurrence.

The second category refers to the strength of a causal generalization (I.1.2). Walton (1996) suggests the critical question: "How strong is the causal generalization (if it is true at all)?" (p. 74). He probably tries to capture two requirements, given his argument

scheme for the argument from cause to effect<sup>53</sup>. First, he appears to demand that the following premise is really true: “Generally, if *A* occurs, then *B* will (or might) occur” (p. 73). Moreover, he wants this generalization to be strong. As he explains about his scheme, “The bridging principle or warrant in the major premise can be variable in strength. In a strong attribution of causality, it might be said that if *A* occurs, then *B* will definitely occur. In a weaker form, it might be said that if *A* occurs, then there is a danger that *B* might occur” (1996, p. 73). Thus, he wants the causal generalization to be true and moreover, he wants it to be invariable. Since the strength of the causal generalization may be used as a way to determine *B*’s probability on the basis of *A*, I consider it a subcategory under (I.1).

Table 3.2 Evaluation questions falling under the category of (I.1) Cause sufficiency criterion.

<i>Evaluation question</i>	<i>Source</i>
I.1 Cause sufficiency criterion	
Is cause <i>A</i> in general sufficient to consider effect <i>B</i> probable? <sup>54</sup>	Schellens (1985, p. 100); Schellens & Verhoeven (1994, p. 82)
Does that what is proposed as a cause ( <i>Z</i> ) indeed lead to the effect mentioned ( <i>Y</i> )? <sup>55</sup>	Garssen (1997, p. 20)
Are there any other factors that, together with what is proposed as a cause ( <i>Z</i> ), should be present to cause the effect mentioned ( <i>Y</i> )? <sup>56</sup>	Garssen (1997, p. 20)
<i>In terms of capability:</i>	
Is the cause capable of producing the desired effect?	Warnick & Inch (1989, p. 109)
Does the cause carry enough force to produce the effect?	Reinard (1991, p. 199)
Is the cause capable of producing the effect?	Freeley & Steinberg (2000, p. 162)

<sup>53</sup> This is the scheme presented:  
 Generally, if *A* occurs, then *B* will (or might) occur.  
 In this case, *A* occurs (or might occur).  
 Therefore, in this case, *B* will occur (or might occur) (1996, p. 73).

<sup>54</sup> In Dutch: “Is oorzaak *A* in het algemeen voldoende om gevolg *B* waarschijnlijk te achten?” (1994, p. 82). In Schellens (1985, p. 100), the formulation is slightly different: “Is cause *A* in general sufficient to consider effect *B* certain or probable?”. In Dutch: “Is oorzaak *A* in het algemeen voldoende om gevolg *B* zeker of waarschijnlijk te achten?”

<sup>55</sup> In Dutch: “Leidt datgene wat als oorzaak wordt voorgesteld (*Z*) inderdaad tot het genoemde gevolg (*Y*)?” (1997, p. 20).

<sup>56</sup> In Dutch: “Zijn er nog andere factoren die samen met datgene wat als oorzaak wordt voorgesteld (*Z*) aanwezig moeten zijn om het genoemde gevolg (*Y*) te doen optreden?” (1997, p. 20).

I.1.1 Contributing and/or counteracting factors criterion	
Might some other cause offset the desired effect?	Warnick & Inch (1989, p. 110)
Are other conditions inhibiting the causal relationship?	Reinard (1991, p. 199)
Are there any circumstances in this particular case that decrease (or increase) the chance of B? <sup>57</sup>	Schellens (1985, p. 100); Schellens & Verhoeven (1994, p. 82)
Are there other factors that would or will interfere with or counteract the production of the effect in this case?	Walton (1996, p. 74)
Is there a counteracting cause?	Freeley & Steinberg (2000, p. 161)
I.1.2 Strength of the causal generalization criterion	
How strong is the causal generalization (if it is true at all)?	Walton (1996, p. 74)

### (I.2) Cause accuracy criterion

Schellens and Verhoeven formulated the question: “Are there any reasons to doubt fact A?” (1994, p. 82, translated from Dutch, see also Table 3.3). That question can be traced back to the more general acceptability criterion. According to the acceptability requirement, “the premises of an argument should be acceptable to the arguer, the audience to whom the argument is directed, and generally to the critical community in which they are situated” (Van Eemeren, Grootendorst & Snoeck Henkemans, 1996, p. 179). However, as the question is formulated so specifically that it applies only to the argument scheme underlying the argument from cause to effect, I consider it a question belonging to the category ‘specific to the argument from cause to effect’ (I).

It should be noted that this question is not relevant to the variant of the argument from cause to effect, in which the prediction is conditional. After all, the cause is not a fact in this variant that should be doubted, but a cause that may or may not occur in future (cf. Schellens, 1985, p. 100).

---

<sup>57</sup> In Dutch: “Zijn er in dit concrete geval nog omstandigheden die de kans op B verkleinen (of vergroten)?” (1994, p. 82). In Schellens (1985, p. 100), it says “Bi” instead of “B”.

Table 3.3 Evaluation questions falling under the category of (I.2) Cause accuracy criterion and (I.3) Cause relevance criterion.

<i>Evaluation question</i>	<i>Source</i>
I.2 Cause accuracy criterion	
Are there any reasons to doubt fact A? <sup>58</sup>	Schellens (1985, p. 100); Schellens & Verhoeven (1994, p. 82)
I.3 Cause relevance criterion	
Does fact A connect adequately to the applied causal relation? <sup>59</sup>	Schellens (1985, p. 100); Schellens & Verhoeven (1994, p. 82)

### (I.3) Cause relevance criterion

Schellens and Verhoeven formulated the question “Does fact A connect adequately to the applied causal relation?” (1994, p. 82, translated from Dutch, see also Table 3.3). This appears to be a specification of a more general relevance criterion. After all, if fact A does not connect adequately to the applied causal relation, then the argument lacks a (clear) relation to the matter at hand. However, just like the question in category (I.2), the relevance criterion is given a specific interpretation and therefore belongs to the category ‘specific to the argument from cause to effect’ (I).

## (II) Generally applicable to all argument schemes

### (II.1) Backup evidence criterion

As Table 3.4 shows, Walton directly appeals to the arguer’s responsibility for producing evidence: “The second critical question asks for evidence cited, if any, to back up this claim” (1996, p. 75). I consider this a general question, because it is applicable to all argumentation schemes.

Table 3.4. Evaluation questions falling under the category of (II.1) Backup evidence criterion and (II.2) Strength of conclusion criterion.

<i>Evaluation question</i>	<i>Source</i>
II.1 Backup evidence criterion	
Is the evidence cited (if there is any) strong enough to warrant the generalization as stated?	Walton (1996, p. 74)
II.2 Strength of conclusion criterion	
Is the probability or certainty expressed in the	Schellens (1985, p. 100);

<sup>58</sup> In Dutch: “Zijn er redenen om het gegeven A te betwijfelen?” (1994, p. 82). In Schellens (1985, p. 100), it says ‘Ai’ instead of ‘A’.

<sup>59</sup> In Dutch: “Sluit het gegeven A goed aan bij het gehanteerde causale verband?” (1994, p. 82). In Schellens (1985, p. 100), the formulation is different: “Is Ai a clear case of A?”. In Dutch: “Is Ai een duidelijk geval van A?”

conclusion in accordance with the answer on question 1-4? <sup>60</sup>	Schellens & Verhoeven (1994, p. 82)
---	-------------------------------------

### (II.2) Strength of conclusion criterion

Schellens (1985, p. 100) and Schellens and Verhoeven (1994, p. 82) present the question: "Is the probability or certainty expressed in the conclusion in accordance with the answer on question 1-4?" (translated from Dutch, see also Table 3.4). They appear to warn for the danger of formulating the conclusion as if it were beyond doubt, for instance: "Casual drug use usually leads to addiction. Students are known for casual drug use. Therefore, it is absolutely guaranteed that they will become addicts". Even if the argument from cause to effect meets all criteria, the conclusion may be formulated in such absolute terms that it is not acceptable. As this criterion can be applied to each conclusion regardless of the type of argument it is supported by, I consider this question to be of a more general nature.

### (III) Specific to different argument schemes

#### (III.1) Both cause necessity and cause sufficiency criterion

According to Warnick and Inch (1989), Reinard (1991) and Freeley and Steinberg (2000), the cause should be both necessary and sufficient for the effect (see Table 3.5). That means that one should question whether the non-occurrence of cause *A* guarantees the non-occurrence of effect *B* (necessity) *and* whether the occurrence of cause *A* brings about the occurrence of effect *B* (sufficiency). It can be argued that questioning cause necessity is only relevant to the argument from *effect to cause* and not to the argument from *cause to effect*. After all, in order to predict an effect it is enough to confirm that the proposed cause is a sufficient cause (cf. Schellens, 1985, pp. 96-97).

Table 3.5 Evaluation questions falling under the category of (III.1) Both cause necessity and cause sufficiency criterion.

<i>Evaluation question</i>	<i>Source</i>
III.1 Both cause necessity and cause sufficiency criterion	
Is the cause necessary and sufficient?	Warnick & Inch (1989, p. 109)
Is there a necessary and sufficient cause for the effect?	Reinard (1991, p. 199)

<sup>60</sup> In Dutch: "Is de in de conclusie uitgedrukte waarschijnlijkheid of zekerheid in overeenstemming met het antwoord op vraag 1-4?" (1985, p. 100; 1994, p. 82).



Is the cause necessary and sufficient?	Freeley & Steinberg (2000, p. 162)
--	------------------------------------

### (III.2) Cause necessity criterion

Reinard (1991), Freeley and Steinberg (2000), and Meany and Shuster (2002) formulated questions that refer to the cause as a necessary condition (see Table 3.6). Reinard (1991, p. 199) asks: “Are other causes relevant in explaining effects?” Then he explains: “There usually are many causes of a problem. Hence, we usually are skeptical of arguments that assume a single cause for a given effect”. So, he implies that one should look for alternative causes that may have preceded the effect. In other words: one should wonder if the supposed cause is necessary for the effect. Meany and Shuster (2002) suggest something similar: “Are there other causes that could have prompted the discussed effect?” (p. 62).

Freeley and Steinberg (2000, p. 160) also aim at questioning the necessity of the cause: “Is this the sole or distinguishing causal factor?”. This question is clarified as follows: “Advocates should determine whether the alleged cause is the only causal factor in producing the effect under consideration or, if not, whether it is the distinguishing causal factor”. So, again, this appears to be another way of asking if the supposed cause is necessary for the effect. As stated above for category (III.1), questioning cause necessity appears to be inappropriate for this particular argument type.

Table 3.6 Evaluation questions falling under the category of (III.2) Cause necessity criterion.

<i>Evaluation question</i>	<i>Source</i>
III.2 Cause necessity criterion	
Are other causes relevant in explaining effects?	Reinard (1991, p. 199)
Is this the sole or distinguishing causal factor?	Freeley & Steinberg (2000, p. 160)
Are there other causes that could have prompted the discussed effect?	Meany & Shuster (2002, p. 62)

### (III.3) Association or causality criterion

The questions in this category are formulated by Reinard (1991, p. 199) and Freeley and Steinberg (2000, p. 160): “Are the causes and effects related in a direct way?” and “Is the alleged cause relevant to the effect described?” (see Table 3.7). Although the questions differ in formulation, they appear to warn that association does not necessarily imply a causal relationship. In other words: they appear to

warn of a *post hoc ergo propter hoc*<sup>61</sup>. After all, Reinard (1991) explains: “A causal relationship claims that there is some reason to think that one element influences the behavior of the other (...) We should be very careful about drawing causal relationships between things that are coincidentally associated” (p. 199).

Freeley and Steinberg (2000) too, say that one should not infer a causal relationship when two events follow each other, as is the case in superstitions: “The superstition that breaking a mirror will cause seven years of bad luck, for example, is based on the assumption that a cause-to-effect relationship exists when in fact there is no such relationship. Unless and until a causal link can be established between an alleged cause and an alleged effect, one cannot hope to develop causal reasoning” (p. 160).

Based on a desire for people to distinguish causal from non-causal relations, they actually refer to a scheme in which our reasoning takes us from a positive correlation between *A* and *B* to a causal connection<sup>62</sup>. The fundamental difference between the scheme, in which it is reasoned from correlation to cause, and the scheme from cause to effect, is that in the first scheme, it is reasoned from non-causal regularity whereas in the second, it is reasoned from causal regularity (Schellens, 1985; Schellens & Verhoeven, 1994). So, the questions of Freeley and Steinberg (2000), and also Reinard (1991), are specific to another scheme than the one underlying the argument from cause to effect.

Table 3.7 Evaluation questions falling under the category of (III.3) Association or causality criterion, (III.4) Nature of side-effects criterion and (III.5) Counterplan criterion.

<i>Evaluation question</i>	<i>Source</i>
III.3 Association or causality criterion	
Is the alleged cause relevant to the effect described?	Freeley & Steinberg (2000, p. 160)
Are the causes and effects related in a direct way?	Reinard (1991, p. 199)

<sup>61</sup> This fallacy is committed when “a cause-effect relation is inferred from the mere observation that two events take place one after the other” (Van Eemeren, Grootendorst & Snoeck Henkemans, 1996, p. 302).

<sup>62</sup> Walton (1996), for instance, presents a separate scheme underlying the argument from correlation to cause  
 “There is a positive correlation between *A* and *B*. Therefore, *A* causes *B*” (p. 71).  
 Schellens and Verhoeven (1994) present a similar scheme: “*A* is the case. *B* is the case. Therefore: *A* is probably the effect (or the cause) of *B*” (p. 85, translated from Dutch).

III.4 Nature of side-effects criterion	
Is there reasonable probability that no undesirable effect may result from this particular cause?	Freeley & Steinberg (2000, p. 161)
Have important countervailing effects been ignored?	Reinard (1991, p. 199)
What other effects does the cause produce? How do these weigh against the already specified causes? <sup>63</sup>	Meany & Shuster (2002, p. 62)
III.5 Counterplan criterion	
How does a new cause affect the system?	Freeley & Steinberg (2000, p. 163)

#### (III.4) Nature of side-effects criterion

As Table 3.7 shows, Freeley and Steinberg (2000) present the question: "Is there reasonable probability that no undesirable effect may result from this particular cause?" Then they explain: "Usually a given cause will produce various effects in addition to the effect under consideration. Will these other effects be desirable, unimportant, or undesirable?" (p. 161). Reinard's (1991) question is: "Have important countervailing effects been ignored?" As he explains, countervailing effects are "other effects" producing "results in a contrary direction" (p. 199). Meany and Shuster (2002) are in line with the other authors with their suggestion to ask how other effects weigh up against the ones already specified<sup>64</sup>.

There appears to be a subtle difference between Freeley and Steinberg (2000) on the one hand and Reinard (1991) on the other hand. From their presented examples, it can be inferred that Reinard (1991) only aims at the nature of side-effects that are in a similar dimension as the effect under consideration (reducing capital gains taxes may stimulate economic growth, which is a desirable consequence, but may also increase inflation, which is an undesirable side-effect in the same economical dimension). Freeley and Steinberg's (2000) criterion is less restricted: the undesirable side-effect may be in the same dimension as the desirable effect under consideration (penicillin may cure a certain illness and thus improve health, but the side-effects of penicillin may be damaging to health). However, the undesirable side-effects may also be in a different dimension than the desirable effect under consideration (banning television commercials for sugar-laden cereals may improve

<sup>63</sup> It is very likely that the authors in fact meant 'against the already specified effects' instead of "against the already specified causes" (2002, p. 62). The question is presumably about (negative) effects that counterbalance the (positive) effects that the protagonist already specified in the argumentation.

<sup>64</sup> See the previous note.

childrens' health, which is a desirable effect in the health-dimension, but may also result in unemployment among sugar producers, which is an undesirable effect in an economical dimension).

In judging the way an effect is predicted, it should be irrelevant if additional effects are (un)desirable, as long as the effect under consideration will be produced. Rather, this criterion applies to pragmatic reasoning, in which positive and negative consequences of an action are weighted.

#### (III.5) Counterplan criterion

Freeley and Steinberg (2000) formulate the question: "How does a new cause affect the system?" (p. 163, see also Table 3.7). They present the example of people claiming that better medical care is needed for slum residents and others arguing that instead of sending residents back to the slums after being treated, one should invest in providing better housing, better food etcetera. With this example, Freeley and Steinberg (2000) imply that one should look for better actions than the ones proposed. As a consequence, their question appears more connected to pragmatic argumentation than to the argument from cause to effect (cf. Schellens, 1985, p. 97).

### **3.2.3 Conclusion and discussion**

In argumentation theory, critical questions have been formulated that intend to evaluate the argument from cause to effect. Critical questions for causal reasoning in general - i.e. for the argument from cause to effect as well as for the argument from effect to cause - have also been formulated. These questions can be classified according to (I) questions specific to the argument from cause to effect, (II) questions that are generally applicable and (III) questions specific to different argument schemes. Within the first category (I), three subcategories can be distinguished: (I.1) cause sufficiency criterion, (I.2) cause accuracy criterion and (I.3) cause relevance criterion. Within the second category (II), two subcategories can be distinguished: (II.1) backup evidence criterion and (II.2) strength of conclusion criterion. Within the third category (III), a distinction is made between five subgroups of questions: (III.1) both cause necessity and cause sufficiency criterion, (III.2) cause necessity criterion, (III.3) association or causation criterion, (III.4) nature of side-effects criterion and (III.5) counterplan criterion.

It can be argued that similar to the cause relevance criterion, as suggested by Schellens (1985) and Schellens and Verhoeven

(1994), there should also be an *effect* relevance criterion. It sounds reasonable to demand that not only the cause in the conclusion (Ai) is a clear member of the kind of causes (A) presented in the argument, but also that the effect in the conclusion (Bi) is a clear member of the kind of effects (B) presented in the argument. For the argument from cause to effect, the relevance criterion should apply to both cause and effect.

To what extent are the contemporary philosophical theories of causation reflected in the categories of evaluation questions? As Table 3.8 shows, a parallel can be drawn between the philosophical concepts of causality, as treated in 3.1.1 and argumentation-theoretical ideas about the evaluation of the argument from cause to effect.

Table 3.8 Comparison between philosophical and argumentation-theoretical views on causality conditions.

<i>Philosophy (Hulswit, 2002)</i>	<i>Argumentation theory</i>
(I) Necessary and/or sufficient conditions approach	(III.1) Both cause necessity and cause sufficiency criterion (III.2) Cause necessity criterion (I.1) Cause sufficiency criterion
(II) Counterfactual approach	(III.2) Cause necessity criterion
(III) Instrumental/agency approach	(I.1) Cause sufficiency criterion <i>in terms of capability</i>
(IV) Probabilistic approach	(I.1.1) Contributing and/or counteracting factors criterion (I.1.2) Strength of the causal generalization criterion

First, the necessary and/or sufficient conditions approach to causality (I) is reflected in the evaluation questions in the categories I.1 and III.1-2. Whether the cause should be a necessary or sufficient condition for the effect or both, these criteria are in line with the philosophical necessary and/or sufficient conditions approach.

Second, the evaluation questions that fall into the category 'cause necessity criterion' (III.2) also reflect the idea of counterfactual dependence (II): if *A* had not happened, *B* would not exist. After all, "a cause is the necessary condition [for] its effect, in the sense that the effect could not have occurred but for the occurrence of the cause. Saying that the short circuit was a necessary condition [for] the fire, is

saying that, given the circumstances, this fire could not have occurred without this short circuit" (Hulswit, 2002, p. 54).

Third, argumentation theorists require that a cause is capable of producing its effect (I.1). Such a requirement fits the instrumental or agency approach of causality (III), in which a cause is an event that we can control in order to produce a desired event or to prevent an undesired event. The instrumental perspective on causality is also inherent in pragmatic argumentation, in which an act is promoted by saying that it is a way to bring about its desired effect.

Finally, the idea that contributing and/or counteracting factors (I.1.1) influence the quality of the argument from cause to effect corresponds to the probabilistic approach to causality (IV), in which it is assumed that events might change the probability of other events. These modifying factors can be contributing, making the effect more likely, or counteracting, making the effect less likely.

Walton's (1996) question about the strength of the causal generalization (I.1.2) also appears to match up with the probabilistic approach to causality (IV). After all, if a causal relationship remains constant under conditions, then apparently the probability of the occurrence of B is higher than when the relationship does not always hold.

### **3.3 Laymen criteria**

In section 3.2, I reported argumentation-theoretical conceptions of the quality of the argument from cause to effect. In this section, the central question is: what criteria are used by laymen to evaluate the quality of the argument from cause to effect? Next, the method to address this question will be discussed.

#### **3.3.1 Method**

##### **Material**

The material consisted of an open case and a closed case. Individual interviews, in which respondents are stimulated by an open case and a closed case to reflect on norms for argument quality, appeared to be the most effective method to uncover laymen criteria (see chapter 2 and Šorm, Timmers & Schellens, 2007). Therefore, this method was used again.

*Open case (Appendix 3A-B).* In the open case, respondents were asked to imagine themselves in a job interview, applying for a job as a speechwriter for a particular ministry (the name of the ministry was dependent on the claim). Within the interview, they received an assignment to test whether they could write good speeches. The assignment was to write down one strong and one weak argument supporting a claim in one of the minister's speeches. In one version, the claim was (1) "If the maximum permitted fat content in minced meat is reduced from 35 to 25 percent, the costs of health care will decrease" (In Dutch: "Als we het maximaal toegestane vetgehalte in gehakt van 35 naar 25 procent verlagen, dan zullen de kosten voor de gezondheidszorg dalen"). In the second version, the claim was (2): "If there is more greenery in and surrounding new housing estates, residents of these estates will use their cars less" (In Dutch: "Als er meer groen in en rond nieuwbouwwijken komt, dan zullen de bewoners van die wijken minder gebruik maken van de auto"). The respondents were also told that the interviewer would ask them why one argument was stronger than the other.

To increase the chance of the respondents coming up with the intended argument type, the respondents were told that they 'had decided to support the minister's claim with predictions'. Moreover, the respondents were to complete a sentence directing them to the argument type I was interested in: "Because if we lower the fat content in minced meat, then.... /Because if the new housing estates are greener, then...". (In Dutch: "Want als we het vetgehalte in gehakt verlagen, dan.../Want als de nieuwbouwwijken groener worden, dan...").

After the respondents had written down their arguments, the interviewer asked them to read their arguments aloud and to explain why one argument was stronger than the other. The explanation allowed for identification of the respondents' evaluation criteria.

Putting a respondent in the position of a speech writer during a job assessment created a situation where it was natural to come up both with a strong and a weak argument in favor of a typical policy claim, as well as to explain the differences between these arguments.

*Closed case (Appendix 3C-D).* The closed case consisted of a claim and five supporting arguments. The claim was either (3) "If employers give employees who are absent due to long-term illness the opportunity to telework, they will return to their regular workplaces sooner" (In Dutch: "Als werkgevers langdurig zieke medewerkers in de gelegenheid stellen om te telewerken, dan zullen ze eerder

terugkeren naar hun gewone werkplek”) or (4) “Giving more money to youth drama schools will promote livability in society” (In Dutch: “Als er meer geld beschikbaar wordt gesteld voor jeugdtheaterscholen, dan zal de leefbaarheid van de samenleving bevorderd worden”). Each claim was introduced with a few sentences about the claim’s subject matter, such as a definition of telework or what happens in youth drama schools.

The arguments were set up according to two schemes: (1) an argumentation scheme based on a few conditional predictions and (2) an argumentation scheme based on a general prediction. Schematically (Schellens, 1985, p. 92, translated from Dutch; see also 3.1.2):

- (1) If  $A_i$  then (probably)  $B_i$   
 If  $B_i$  then (probably)  $C_i$   
 Therefore: if  $A_i$  then (probably)  $C_i$
- (2) A generally leads to B  
 Therefore: if  $A_i$ , then probably  $B_i$

It should be noted that when scheme (1) was used, the number of conditionals in the argument varied. Hence, in fact, variants of argument scheme (1) were used for the closed case.

Two of the five arguments were supposed to be strong. Three of the five arguments were considered weak, as they violated certain criteria suggested in argumentation theory (see also section 3.2). The criteria I aimed to violate were (1) the cause sufficiency criterion (‘*This* should be enough to cause *that*’), (2) the cause relevance criterion (‘The cause in the argument should connect adequately to the cause in the conclusion’) and (3) the effect relevance criterion (‘The effect in the argument should connect adequately to the effect in the conclusion’). Although the distinction between criteria (2) and (3) has not been made in theory and only the cause relevance criterion was revealed by the literature study, it can be argued that the relevance criterion should be applied to both cause *and* effect (see also 3.2.3). So, I wanted to know whether in practice people pay attention to relevance and if they do, whether they focus on the relevance of the cause, on the relevance of the effect or on both. I applied criterion (1) only on scheme (1) and criterion (2) and (3) only on scheme (2). Table 3.9 gives an overview of the evaluation criteria that the arguments were supposed to violate.



Table 3.9 Evaluation criteria violated in the closed-case arguments (translated from Dutch).

		<i>Claim (3)</i>	<i>Claim (4)</i>
		If employers give employees who are absent due to long-term illness the opportunity to telework, they will return to their regular workplaces sooner.	Giving more money to youth drama schools will promote livability in society.
<i>scheme; criterion violated</i>		<i>Argument</i>	<i>Argument</i>
scheme (2); weak, violates cause relevance criterion	A	Employees who have been ill for a long time and who get a holiday will generally return sooner to their regular workplace.	Initiatives that involve people in society generally lead to a more livable society.
scheme (1); weak, violates sufficiency criterion	B	If employees who have been ill for a long time are able to telework, then they will be tired less quickly. Eventually they will return to their regular workplace sooner.	If youth drama schools get a financial impulse, then young people will be better educated in amateur theatre. Eventually, society will be more livable.
scheme (2); weak, violates effect relevance criterion	C	Measures to let employees who have been ill for a long time work from home generally lead to the solution of social and individual problems.	Investing in youth drama schools generally leads to an enduring future.
scheme (1); strong, no criteria violated	D	If ill employees can do their work from home behind the computer, then they won't have to travel to work. And if they won't have to travel to work, then they will be less burdened. And if they will be less burdened by traveling, then they will soon be able to take on a bit of work.	If youth drama schools get financial support, then they can strengthen themselves artistically and organizationally. And if youth drama schools are qualitatively stronger, then children and young people will be better educated there. And if they are better educated in the domain of arts and culture, then they will generally develop better as well.
scheme (1); strong, no criteria violated	E	If ill employees can telework, then they will take on more work. Moreover, they can stay involved with work and colleagues. Eventually they will return to their regular workplace sooner.	If more money is put at the disposal of youth drama schools, more young people will get the chance to become involved in society. Moreover, young people will educate themselves better. Eventually, society will become more livable.

The respondents were asked to arrange the arguments in order of strength, with the strongest argument in first place and the weakest in fifth. After the respondents had ranked their arguments, the interviewer asked why a particular argument was put first, why another argument was put second and so on. Criteria were derived on the basis of their explanations.

### **Respondents**

Twenty respondents were interviewed. All respondents were visitors to a public library in Nijmegen, the Netherlands. Each of the respondents was paid ten Euros for participating. Of the respondents, eleven were male and nine were female. The ages of the respondents varied from 17 to 65, and the average age was 36. The level of education varied from university education ('WO') to preparatory vocational education ('VMBO'). The majority (11) had received higher professional education ('HBO') and university education ('WO'). None of the respondents had received any formal training in the field of argumentation theory.

### **Booklet**

All respondents received a booklet. Each booklet consisted of five pages. On the first page, an instruction text was printed. It stated that the research was about 'government communication' and that there were no wrong answers. The second page contained the open case, and the third page contained the closed case. The fourth page contained some general questions about the respondents' level of education, sex and age. The respondents also had the chance to write down what they thought was the purpose of the study.

On the last page of the booklet, questions were asked to test whether the respondents had received any formal training in the field of argumentation theory. The first question was if they had ever been in contact with argumentation studies<sup>65</sup>. If so, they were requested to answer the question if they had ever heard of argument types, if they had ever heard of argumentation schemes and if they had ever heard of ways to evaluate or test arguments. If they answered the last three questions positively, they were invited to write down what the concept

---

<sup>65</sup> I acknowledge that in English as well as in Dutch ("Bent u wel eens in aanraking gekomen met argumentatieleer?"), the formulation is awkward. A better alternative would have probably been to ask if someone had ever been occupied with/engaged in argumentation studies.

was (e.g. 'an argument type is...') and to give an example of the concept (e.g. 'an example of an argument type is...').

## Design

Two open cases and two closed cases were created. Each case revolved around a different claim so that the criteria would not be claim-dependent. The open case was always presented first, to make sure that the responses would not be influenced by the reactions in the closed case. This design led to four different combinations:

- 1) open case/ claim 1; closed case/ claim 3 ( $n = 5$ )
- 2) open case/ claim 1; closed case/ claim 4 ( $n = 5$ )
- 3) open case/ claim 2; closed case/ claim 3 ( $n = 5$ )
- 4) open case/ claim 2; closed case/ claim 4 ( $n = 5$ )

## Procedure

The respondents were personally approached by the interviewer at the entrance of the public library. They were asked to participate in an interview that would take approximately 30 minutes and that would be rewarded with ten Euros. Visitors willing to participate were brought to a separate room in the library. Before the actual interview started, the interviewer informed the respondents that the topic of the interview would be 'government communication'. In addition, they were told that there were no 'wrong' answers.

After the introduction, the open case was presented. When the respondents had mentioned the arguments, they were asked why one argument was stronger than the other. Subsequently the closed case was presented. When the respondents had arranged the arguments in order of strength, they were asked why they had put a certain argument first, second and so on. The interviewer used some techniques to get more information, without directing the respondents. She asked questions like 'Can you explain that?' and 'What/how do you mean?'.

After the actual interview, the respondents were asked whether they had any comments on the interview. Afterwards, they filled in personal details (age, sex and level of education), an answer to the question as to what they thought the purpose of the study was and answers to questions regarding their knowledge of argumentation theory.

Finally, they received a 'thank you' for their cooperation and were paid ten Euros. On average, it took approximately 30 minutes to

complete an interview (including introduction and additional questions). All interviews were recorded on tape and transcribed afterwards.

### Data-analysis

To analyze the data, I used the same technique as used in the study on the argument from authority: a content analysis (see chapter 2). Here, I will only present an outline of the data-analysis performed for this study and give some examples for illustration purposes. As the performance of the content analysis differed slightly from the one performed for the argument from authority, I will also indicate these differences. The data, consisting of transcripts, were analyzed according to the following steps:

(1) Selecting relevant data and deleting irrelevant data, such as utterances related to the task instruction: 'Eh...well I think, eh. Can I read it again?' or 'So eh this was the first case, let's go on to the second'. (In Dutch: 'Eh... nou ik denk, eh. Mag ik hem nog een keer doorlezen?' or 'Dan eh was dat de eerste casus, dan gaan we door naar de tweede').

(2) Dividing relevant text into fragments, on the conditions that one fragment would be about one topic (e.g. the evaluation of one particular argument), comprehensible and not taken out of context. For example:

- (1) Yes, er, I put E on 1, because it says, then more young people will get the chance to be involved in society. And I think that it is simply important that more people are involved in society.

(In Dutch:

Ja, eh, E heb ik op 1 gezet, omdat er staat, dan krijgen meer jongeren de kans om betrokken te worden bij de maatschappij. En ik denk dat het gewoon belangrijk is dat meer jongeren betrokken worden bij de maatschappij.)

(Argument E was:

If more money is put at the disposal of youth drama schools, more young people will get the chance to become involved in society. Moreover, young people will educate themselves better. Eventually, society will be more livable.)

- (2) I chose A, because initiatives that involve people in society, so that that always leads to a er better, more livable society er becomes (...). I find this a strong argument, because er because it is about

er initiative, that people should take initiative. And if many people take initiative and if they are on the same level, then you get a better, or a more livable society.

(In Dutch:

Ik heb A gekozen, omdat initiatieven die mensen bij de samenleving betrekken, dat dus altijd dat dat tot een eh betere, leefbaardere samenleving eh wordt (...). Ik vind hier sterk aan, omdat eh omdat het dus gaat over eh dat initiatieven, dat mensen initiatieven moeten nemen. En als veel mensen initiatief moeten nemen en veel neuzen dezelfde kant opstaan, dan krijg je gewoon een betere, of een leefbaardere samenleving.)

(Argument A was:

Initiatives that involve people in society generally lead to a more livable society.)

These are considered two separate fragments, because each fragment deals with a different argument ('A' and 'E').

(3) Attaching criteria to the fragments. For the argument from authority, we defined a criterion as a norm for evaluating an argument that applies to every argument from authority (see also chapter 2). If I had used this definition, it would have caused problems in the analysis, because most respondents stayed close to the particular context in their verbalizations. In general, they did not use abstract phrases such as 'This proposed action is enough to cause that consequence', but rather "And if many people take initiative and if they are on the same level, then you get a better, or a more livable society" (see example 2 above). I assumed that more abstract norms underly these specific phrasings. To be able to identify and classify these abstract norms, I assigned symbols to the measures (M) and effects (G, I, J, E<sup>66</sup>) in the material. Then I assigned the same symbols to corresponding measures (M) and effects (G, I, J, E) in the respondents' utterances. For instance, I attached the following criteria to the fragments above:

- (3) Is it important to involve more young people in society (G)?/Is het belangrijk om meer jongeren te betrekken bij de maatschappij (G)?

(Argument E was: If more money is put at the disposal of youth drama schools (M), more young people will get the chance to

---

<sup>66</sup> The closed case material contained arguments with more than one effect. For these arguments, I used a different letter for each different effect. So, for an argument in which four different effects were mentioned, I used the four different letters G, I, J, and E.

become involved in society (G). Moreover, young people will educate themselves better (I). Eventually, society will be more livable (E.)

- (4) If more people take initiative (M), do you get a more livable society (E)?/Als veel mensen initiatieven nemen (M), krijg je dan een leefbaardere samenleving (E)?

(Argument A was: Initiatives that involve people (M) in society generally lead to a more livable society (E).)

Assigning the symbols to elements within the utterances made it easier to abstract from the specific formulations of the respondents. In case of fragment (1) presented above, the abstract criterion became “Is the effect (G) important?” In case of fragment (2), the abstract criterion became: “If measure (M), then (E)?”

(4) Categorizing the questions that were registered in the previous analysis step on the basis of a shared characteristic. The questions in (3) and (4), for instance, belonged to two different categories, as the former refers to the desirability of the effect and the latter to cause sufficiency. I provided each category with a label in the form of one or more catchwords (e.g. ‘desirability effect’ or ‘cause sufficiency’).

### **Interrater agreement**

To assess the extent to which fragments could be reliably related to criteria, interrater agreement was determined. The second rater was a colleague working at the Department of Business Communication Studies of the Radboud University Nijmegen (the Netherlands), who has special knowledge in the field of argumentation. First, she received a written task instruction. It stated that she was to judge a number of fragments, in which phrases were marked. In addition, it explained that her task would be to label the phrases with criteria. The rater received a list of possible criteria to choose from. Each criterion on the list was defined. The rater was also asked to study a document with clarifications of the terms ‘necessary condition’, ‘sufficient condition’ and ‘necessary and sufficient condition’, to make sure she would not have any troubles understanding these concepts in the criteria definitions. Furthermore, the rater received the closed cases used in the interviews.

Five training fragments were presented, to give the second rater the opportunity to familiarize herself with the instrument and to ask for clarification. Afterwards, 10 randomly selected test fragments

were presented. The results showed that the Kappa Measure of Agreement value was .636 with a significance of  $p < .0005$  ( $N = 12$ ). According to Peat and Barton (2005, p. 268), this value generally represents good agreement.

### 3.3.2 Results

Next, I will discuss the laymen criteria for evaluating the argument from cause to effect. First, the results of the open case will be reported, followed by the results of the closed case.

#### Interviews open case

It appeared from an analysis of interviews in the open case that all respondents made up arguments with causal sequences. The sequences varied in the number of causal steps. Most respondents made one or two causal steps in one argument. Less frequently, they presented three steps in one argument. Table 3.10 provides an illustration:

Table 3.10 Illustration of respondents' predictions in the open case.

	Sentence to complete in the open case	One prediction	Two predictions	Three predictions
In English:	Because if we lower the fat content in minced meat, then...	the chance of obesity will be reduced (1).	in any case the amount of fat people ingest per meat-containing meal will be lowered (1) and a strong increase of cholesterol level will be reduced (2).	the level of saturated fatty acids will be lower (1) which will lower the chance of cardiovascular diseases (2), and will reduce costs (3).
In Dutch:	Want als we het vetgehalte in gehakt verlagen, dan ....	neemt de kans op overgewicht af (1).	daalt sowieso het vetgehalte wat mensen per vleesmaal binnenkrijgen (1) en daalt ook de sterke toename van cholesterolgehalte (2).	is het gehalte aan verzadigde vetzuren lager (1) en daardoor de kans op hart- en vaatziekten kleiner (2), waardoor de kosten dalen (3).

All respondents succeeded in creating arguments containing causal sequences. However, not all data were informative regarding criteria for evaluating the argument from cause to effect. Some explanations were simply not clear enough to serve as a basis for inferring evaluation criteria, others presented an argument in favor of the action in the claim as a strong argument and an argument against the action in the claim as a weak argument, as illustrated by the following fragment:

Fragment (1)

Claim to be supported:

If there is more greenery in and surrounding new housing estates, residents of these estates will use their cars less.

In English:

I: Er, can you read out loud what the strongest argument is?

R: Er, that civilians become more aware of the environments. And the less, er weaker argument is, yes, there is enough greenery so er, why should we?

In Dutch:

I: Ehm, kunt u voorlezen wat het sterke argument is?

R: Ehh, dat burgers zich meer bewust worden van het milieu. En het minder, ehh zwakke argument is, ja, er is toch genoeg groen dus ehh, waarom zouden we ook?

In addition, it appeared that some respondents made the distinction between the strong and the weak argument on the basis of the desirability of a consequence in the argument. For example:

Fragment (2)

Claim to be supported:

If the maximum permitted fat content in minced meat is reduced from 35 to 25 percent, the costs of health care will decrease.

In English:

R: The strong argument, because if we reduce the fat content in minced meat, then the number of cardiovascular diseases will go down and along with that the costs in health care (...) The weak argument. If we reduce the fat content in minced meat, then the pure taste of minced meat will come out more.

I: Okay, can you explain what er, why one argument is stronger than the other argument?

R: Er, well I think that the politician does not care much about the taste of pure minced meat in any case (...) And he does care about the health care costs (...) So er, that is the strong part of it, I think (...) And er yes, I think that it is a very obvious advantage, if you would do that (...) So the chance, that is cost number one in health care I think, so I think that would yield much profit. And the taste of, that taste of good minced meat does not, that is just pleasant, but that is not of any particular advantage to a politician.



In Dutch:

R: Het sterke argument, want als we het vetgehalte in gehakt verlagen, dan daalt het aantal hart- en vaatziekten en daarmee de kosten in de gezondheidszorg (...)  
Het zwakke argument. Als we het vetgehalte in gehakt verlagen, dan komt de zuivere smaak van gehakt beter naar voren.

I: Oké, kunt u uitleggen wat ehh, waarom het ene argument sterker is dan het andere argument?

R: Ehh, nou ik denk dat de politicus zich sowieso niet zoveel zorgen maakt om de smaak van zuiver gehakt (...) En wel om de kosten van de gezondheidszorg (...) Dus ehm, dat is al het sterke daarvan, denk ik (...) En ehh ja, ik denk dat dat een heel evident voordeel is, als je dat zou doen (...) Dus de kans, dat is toch kostenpost nummer één in de gezondheidszorg denk ik, dus ik denk dat dat veel voordeel op zou leveren. En die smaak van, die smaak van lekker gehakt niet, dat is alleen maar prettig, maar verder heeft dat voor een politicus geen enkel voordeel.

In the open case, the data of six out of twenty respondents were useful, as their norms were comprehensible and clearly related to the argument from cause to effect. Here I present fragments of interviews to illustrate these evaluation criteria. In each fragment, the strong argument is indicated by '(+)', the weak argument by '(-)' and the explanation about why one argument is stronger than the other by '(E)'. The claim that was to be supported by the arguments is also given.

Three of six respondents referred to *the extent to which the causal relation is established*. For instance:

Fragment (3)
--------------

Claim to be supported:

If there is more greenery in and surrounding new housing estates, residents of these estates will use their cars less.

In English:

R: Er if there is more greenery in new housing estates, then people use public transport or bikes more often, because then it is pleasant to cycle through the area.  
(+)

I: Okay, and what did you choose as the weak argument?

R: Er if there is more greenery in the new housing estates, then civilians will use their cars less, because they want to preserve nature and the greenery as much as possible. (-)

I: And why do you think argument A is stronger in supporting the argument, or in supporting the standpoint than argument B?

R: Er, I think that A is somewhat more based on facts. Because I think that if there is more greenery, that they would really er really, you know, go by bike more often and use public transport (...) And er that other one is weaker, I think, because er, yes, you cannot check if civilians want to preserve nature. They may say so, but you cannot completely, that, yes, it is more like, well, an opinion. And that, that other one you can really er, is more of a fact, where you can really check if it is true or not. (E)

In Dutch:

R: Ehh als de nieuwbouwwijken groener worden, dan gebruiken mensen vaker het openbaar vervoer of de fiets, want dan is het aangenaam om door de wijk te rijden. (+)

I: Oké, en wat heb je bij het zwakke argument?

R: Ehh als de nieuwbouwwijken groener worden, dan zullen de burgers minder vaak de auto gebruiken, omdat ze de natuur en het groen zo goed mogelijk willen behouden. (-)

I: En waarom vind je argument A sterker ter ondersteuning van argument, of ter ondersteuning van het standpunt dan argument B?

R: Ehh, ik denk dat A iets meer op feiten is gebaseerd. Want ik denk dat als het groener wordt, dat ze echt ehh echt, zeg maar, vaker de fiets zullen pakken en het openbaar vervoer gebruiken (...) En ehh die andere is denk ik zwakker, omdat ehh ja, je kunt het niet nagaan of de burgers de natuur willen behouden. Dat kunnen ze wel zeggen, maar je kunt niet helemaal, dat, ja, het is eerder een beetje een mening, zeg maar. En die, die andere kun je echt ehh, is meer een feit dat je echt kunt nagaan of het nou klopt of niet. (E)

This respondent is of the opinion that the strong argument is strong because it is a proven principle, whereas the weak argument is weak because the causal relationship cannot be verified.

In three interviews, respondents referred to the *strength of the causal relationship*. Fragment (4) gives an example:

Fragment (4)
--------------

Claim to be supported:

If the maximum permitted fat content in minced meat is reduced from 35 to 25 percent, the costs of health care will decrease.

In English:

R: Yeah, well, the strong argument is er if there is less fat in the minced meat then is, drops the cholesterol level (...). And the people get less heart complaints and that is of course er, then the health care costs will go down of course. (+). And well yeah, otherwise I found it difficult to think of a bad argument, but there I have said, there people become less fat and then, there bones become less burdened (-) (...).

I: And can you explain why you find argument A stronger than argument B? What is the difference between those two?

R: Er well, in any case, that [the proposition in the strong argument] er is therefore a fact, that is, that is an established fact that still (...) But that if you necessarily get fat and your bones are bothering you, that is er, can happen but that does not always have to, so that relation seems less strong to me. (E)

In Dutch:

R: Ja, nou, het sterke argument is ehh als er minder vet in het gehakt zit dan is, daalt het cholesterolgehalte (...) En dan krijgen dus de mensen minder hartklachten en dat is natuurlijk ehh, dan dalen de kosten van de gezondheidszorg natuurlijk. (+) En nou ja, anders vond ik moeilijk om een slecht argument te bedenken, maar daar heb ik gezegd, daar worden de mensen minder dik en dan worden hun botten minder belast (-) (...).

I: En kunt u uitleggen waarom u argument A sterker vindt dan argument B? Wat is het verschil tussen die twee?

R: Ehh nou, in ieder geval, dat [de propositie in het sterke argument] ehh is dus een feit, dat is, dat is een vastgesteld feit dat toch (...) Maar dat als je per se dik wordt dat je last van je botten krijgt, dat is ehh, kan wel voorkomen maar dat hoeft natuurlijk niet altijd, dus dat verband is minder sterk, lijkt mij. (E)

---

This respondent argues that the causal relationship in the strong argument is a verified fact (using a criterion similar to the one mentioned in fragment 3), whereas the causal relationship in the weak argument is less strong than the one in the strong argument. So, in the explanation for what makes the weak argument relatively weak, the respondent refers to the strength of the causal relationship.

In three interviews, respondents referred to *cause sufficiency*.

For instance:

Fragment (5)
--------------

Claim to be supported:

If there is more greenery in and surrounding new housing estates, residents of these estates will use their cars less.

In English:

R: The strong argument, er yes that then there are less er, parking facilities if there is more greenery. (+)

I: Yes, okay. And the weak argument?

R: Er, that then the neighbourhood er becomes a bit more attractive to walk around in, that er you are more tempted to leave your car. (-)

I: Okay, okay, and can you also explain why one argument is stronger than the other?

R: Er, yeah, yeah I think that er, yeah, if there is less, if there is more greenery, that there are indeed slightly less parking facilities. That, yeah, that is more plausible and likely to be the case (...) (E).

In Dutch:

R: Het sterke argument, ehh ja dat er dan minder ehh parkeergelegenheden zijn als er meer groen is. (+)

I: Ja, oké. En het zwakke argument?

R: Ehm, dat de omgeving dan ehh wat aantrekkelijker wordt om in te lopen, dus ehh laat je eerder je auto staan. (-)

I: Oké, oké, en kunt u ook uitleggen waarom het ene argument sterker is dan het andere?

R: Ehh, ja, ja ik denk dat ehh, ja, als er minder, als er meer groen is, dat er dan

---

inderdaad wel wat minder parkeergelegenheden zijn. Dat, ja, dat is meer aannemelijk dat dat zo is (...) (E).

---

The respondent in fragment (5) considers it plausible that if the cause in the strong argument occurs, the effect will follow.

In sum, the interviews on the open case revealed that laymen are concerned with the *extent to which the causal relation is established*, the *strength of the causal relation* and *cause sufficiency*.

### Interviews closed case

The evaluation criteria that were generated by the respondents in the closed case can be grouped into (I) criteria specific to the argument from cause to effect and (II) criteria that are specific to a different argument scheme. Criteria that are not mentioned in theory are in italics.

#### (I) Specific to the argument from cause to effect

##### Cause sufficiency criterion

Nearly all laymen in the current study (95%) use criteria referring to a sufficient condition. It appears that some apply the concept in a strict way and others in a more liberal way. Consider the following fragments:

Fragment (6)
--------------

Evaluated argument (A):

Initiatives that involve people in society generally lead to a more livable society.

Supported claim:

Giving more money to youth drama schools will promote livability in society.

In English:

I: Er, what do you think is strong about that [argument A], or weak?

R: I find this a strong argument, because er because it is about er initiative, that people should take initiative. And if many people take initiative and if they are on the same level, then you get a better, or a more livable society.

In Dutch:

I: Ehm, wat vindt u daar [argument A] sterk aan, of zwak?

R: Ik vind hier sterk aan, omdat eh omdat het dus gaat over eh dat initiatieven, dat mensen initiatieven moeten nemen. En als veel mensen initiatief moeten nemen en veel neuzen dezelfde kant opstaan, dan krijg je gewoon een betere, of een leefbaardere samenleving.

---

Fragment (7)

Evaluated argument (C):

Investing in youth drama schools generally leads to an enduring future.

Supported claim:

Giving more money to youth drama schools will promote livability in society.

In English:

R: Yeah I found that one [argument C] total nonsense. Investing in youth drama schools generally leads to an enduring future. It will contribute, you know, a small part of course, but I don't think that, that it really leads to an enduring future. It will do its share, but, yeah, I did not find it very strong. It is not the case that society all of a sudden completely changes, because more money is put in the art and culture youth drama schools, I think.

In Dutch:

R: Ja die [argument C] vond ik helemaal onzin. Het investeren in jeugdtheaterscholen leidt over het algemeen tot een duurzame toekomst. Het zal wel een klein deel, zeg maar, met zich meebrengen natuurlijk, maar ik denk niet dat, dat het echt leidt tot een duurzame toekomst. Het zal zijn steentje wel bijdragen, maar, ja, ik vond hem niet heel erg sterk. Het is niet dat de samenleving meteen helemaal omslaat, omdat er meer geld in de kunst en cultuur jeugdtheaterscholen wordt gestoken, denk ik.

Fragment (8)

Evaluated argument (B):

If employees who have been ill for a long time are able to telework, then they will be tired less quickly. Eventually they will return to their regular workplace sooner.

Supported claim:

If employers give employees who are absent due to long-term illness the opportunity to telework, they will return to their regular workplaces sooner.

In English:

R: Then I have B. That is actually, yeah it is a bit similar to C. If they can telework, they are tired less quickly. So they have a higher chance of recovery and they will return sooner.

In Dutch:

R: Dan heb ik B. Dat is eigenlijk, ja hij lijkt een beetje op C. Als ze kunnen telewerken, zijn ze minder snel vermoeid. Dus hebben ze meer de kans om te herstellen en keren ze sneller terug.

In fragment (6), the respondent appears to handle the concept of sufficient condition in a strict way: if the circumstance in which many people take initiatives is present, then we get a better society. In other words: if the cause occurs, the effect *must* follow. In fragment (7), the respondent talks in more graded terms: investing in youth drama schools will bring about “a small part” of an enduring future, but it “is

not the case that society all of a sudden completely changes". In other words, the sufficiency of the cause is denied not by the formal indication 'cause, hence no effect', but by the indication 'cause, hence the effect to some extent but not completely'. In fragment (8), the respondent reasons in degrees of likelihood or probability: if the employees can telework, then it raises the probability of recovering and going back to work. So, respondents might treat the concept of sufficiency as stringent as in the formal definition, but they may also use the concept in a more flexible way.

Second, the respondents do not only regard causes as sufficient conditions for effects, but also as instrumental conditions. For example:

Fragment (9)

Evaluated argument (D):

If ill employees can do their work from home behind the computer, then they won't have to travel to work. And if they won't have to travel to work, then they will be less burdened. And if they will be less burdened by traveling, then they will soon be able to take on a bit of work.

Supported claim:

If employers give employees who are absent due to long-term illness the opportunity to telework, they will return to their regular workplaces sooner.

In English:

R: Then I have D. I think that ill employees are ill mainly because they need rest and if they don't need to travel to work, then they become less burdened and they can recover more quickly. And the step to return to work is smaller. So I think that that is also a good reason to do that tele, telework.

In Dutch:

R: Dan heb ik D. Ik denk dat zieke werknemers vooral ziek zijn, omdat ze rust nodig hebben en als ze niet naar hun werk hoeven te reizen, dan worden ze minder belast en kunnen ze sneller genezen. En de stap om dan weer aan het werk te gaan is minder groot. Dus ik denk dat dat ook een goede reden is, om dat tele, telewerken te doen.

Fragment (10)

Evaluated argument (B):

If employees who have been ill for a long time are able to telework, then they will be tired less quickly. Eventually they will return to their regular workplace sooner.

Supported claim:

If employers give employees who are absent due to long-term illness the opportunity to telework, they will return to their regular workplaces sooner.

In English:

I: Right, okay, then B, on the second position...

R: Yes, then they are less tired. If the cause is fatigue, for example caused by long traveling, then this is a solution. Er I don't know if that will lead to them returning sooner, because they still have to travel of course, if they return to work, just as long as before, but perhaps then they have a temporary break to recover from the fatigue, and then they can return to work. So yeah, it's possible, I don't find it a great solution, but it's possible.

In Dutch:

I: Goed, oké, dan B, op de tweede plek...

R: Ja, dan zijn ze minder vermoeid. Als de oorzaak moeheid is, bijvoorbeeld door lang reizen, dan is dit een oplossing. Ehh of ze daardoor eerder terug zullen keren weet ik niet, want het reizen blijft natuurlijk, als ze opnieuw beginnen, weer even lang, maar misschien hebben ze dan tijdelijk een pauze om bij te komen van de vermoeidheid, en komen ze dan weer terug. Dus ja, het kan, ik vind het niet geweldig, maar het kan.

---

In fragment (9), the respondent regards the proposed cause (doing telework) as a “good reason” to establish the chain of events leading to the desired effect (returning to work). In other words: the cause sufficiently serves as a means to an end. In fragment (10), the respondent regards the cause (doing telework) as a possible but not very good solution to a problem (not being able to go to work). So, laymen ask whether they see the cause in the argument as sufficiently instrumental in achieving a goal or solving a problem.

Respondents appear to use certain tests to determine whether a cause is indeed sufficient to bring about an effect. First of all, they take into consideration contributing or counteracting factors: circumstances that might influence the occurrence of the effect. This is illustrated in fragment (11):

Fragment (11)
---------------

Evaluated argument (B):

If employees who have been ill for a long time are able to telework, then they will be tired less quickly. Eventually they will return to their regular workplace sooner.

Supported claim:

If employers give employees who are absent due to long-term illness the opportunity to telework, they will return to their regular workplaces sooner.

In English:

R: But all right, let's see, B, what did I have there? Yeah, the same. If they can telework, they won't get tired as quickly, yeah or tired less quickly compared to the situation in which they have to go to work, that I believe. But that they eventually return sooner to their regular workplace, that has to do with a whole lot of other health-related factors of course (...). Because they should also be cured of er, it totally depends on what he's got. So, so actually also er, you cannot just claim that. Actually not at all.

In Dutch:

R: Maar goed, even kijken, B, wat had ik daarbij? Ja, hetzelfde. Als ze kunnen telewerken, zijn ze minder snel vermoeid, ja of minder snel vermoeid als dat ze naar hun werk zouden moeten gaan, dat geloof ik weer wel. Maar dat ze uiteindelijk dan sneller terug naar hun gewone werkplek, dat heeft met een heleboel andere gezondheidsfactoren te maken natuurlijk (...) Want ze moeten ook genezen zijn van ehh, het hangt er helemaal van af wat hij heeft. Dus, dus eigenlijk ook ehh, kun je niet zo zeggen. Eigenlijk helemaal niet.

---

In fragment (11), the respondent decided that not getting tired as quickly does not necessarily lead to an earlier return to the regular workplace, because there are other health-related factors that may prevent the effect from occurring. So, laymen try to imagine factors that modify the chance that the effect occurs.

Secondly, laymen appear to focus on the strength of the causal generalization in the argument to test whether the cause is enough to bring about the effect. If they think the range of cases the generalization holds for is too narrow, then they doubt that the cause is sufficient in general to produce the effect. This is illustrated in fragment (12):

Fragment (12)
---------------

Evaluated argument (B):

If youth drama schools get a financial impulse, then young people will be better educated in amateur theatre. Eventually, society will be more livable.

Supported claim:

Giving more money to youth drama schools will promote livability in society.

In English:

I: Hmmhmm. Okay, and you've got B on the fourth position?

R: Yeah, if youth drama schools get a financial impulse, then young people will be better educated in amateur theatre. But you first have to become a member of such a school. I don't think that you can say in general terms that young people will be better educated. And that that causes the society to become more livable, that will be caused by the students who, say, actually get that education (...). But there are more students who do not get that education. So it is written here as if the whole society flourishes, because a few students are better educated (...) So I found that a little weird.

In Dutch:

I: Hmmhmm. Oké, en B heb je op de vierde plaats?

R: Ja, als jeugdtheaterscholen een financiële impuls krijgen, dan wordt de jeugd beter opgeleid in amateurtheater. Maar dan moet je eerst wel lid worden van zo'n school dus. Ik denk niet dat je dat in het algemeen kan trekken dat de jeugd dan beter opgeleid wordt (...) En dat de samenleving daar leefbaarder van wordt, dat zal wel zo zijn door de studenten, die zeg maar, daadwerkelijk die opleiding volgen (...). Maar er zijn meer studenten, die die opleiding niet volgen. Dus het staat hier geschreven alsof de hele samenleving opbloeit, omdat een paar studenten beter

---



opgeleid worden (...) Dus dat vond ik een beetje vreemd.

---

As this fragment shows, respondents might have problems accepting the causal generalization. As a consequence, they do not consider the effect certain or probable on the basis of the proposed cause.

Third, laymen in this study appear to judge the causation on the basis of backup evidence or explanation. For instance:

Fragment (13)
---------------

Evaluated argument (B):

If employees who have been ill for a long time are able to telework, then they will be tired less quickly. Eventually they will return to their regular workplace sooner.

Evaluated argument (D):

If ill employees can do their work from home behind the computer, then they won't have to travel to work. And if they won't have to travel to work, then they will be less burdened. And if they will be less burdened by traveling, then they will soon be able to take on a bit of work.

Supported claim:

If employers give employees who are absent due to long-term illness the opportunity to telework, they will return to their regular workplaces sooner.

In English:

R: Then I have B. That is actually, yeah it looks a bit like C. If they can telework, they are tired less quickly. So that means they have a higher chance of recovery and they will return sooner.

I: Hmmhmm, and why do you still think it is a weaker argument than D?

R: Er, because D is more specific about them not having to travel (...) And in B, yeah it's just about them being tired less quickly (...). And that does not necessarily have to be the case, because if they still have to work, they're basically tired just as quickly (...). That's why I think D is better.

In Dutch:

R: Dan heb ik B. Dat is eigenlijk, ja hij lijkt een beetje op C. Als ze kunnen telewerken, zijn ze minder snel vermoeid. Dus hebben ze meer de kans om te herstellen en keren ze sneller terug.

I: Hmmhmm, en waarom vind je hem toch zwakker dan D?

R: Ehm, omdat D specifiek is over dat ze niet hoeven te reizen (...) En bij B, ja gaat het er eigenlijk alleen maar om dat ze minder snel vermoeid zijn (...) En dat hoeft niet per se zo te zijn, want als ze toch ook moeten werken, dan zijn ze in principe net zo snel vermoeid. En bij D gaat het erom waarom ze minder snel vermoeid zijn, omdat ze niet hoeven te reizen (...). Daarom vind ik D beter.

---

Fragment (13) shows that if the causal steps in the chain of events are specified, laymen are more tempted to acknowledge that the effect will follow the cause.

In fragment (14), the respondent requires an explanation for how exactly the cause is supposed to lead to the effect:

Fragment (14)

Evaluated argument (A):

Initiatives that involve people in society generally lead to a more livable society.

Supported claim:

Giving more money to youth drama schools will promote livability in society.

In English:

I: Okay, and A? Because that one is exactly in the middle, in the third position.

R: Er, er. Yeah, I actually thought it had little foundation. So er, yeah it generally led to a more livable society, yeah, why? Yeah, I need to know more as to why. Why do they have that opinion? So that's why I had it a little, yeah, it can be the case, but I found it a little weak (...) Because it was not really further supported.

I: And what do you mean, that it...by supported?

R: Er, well, explanation why. Why, er, why it matters at all, er, causes the society to be more livable anyway and er, in what, in what respect then.

In Dutch:

I: Oké, en A? Want die zit precies in het midden, op de derde plaats.

R: Ehm, ehm. Ja, ik vond hem eigenlijk weinig onderbouwing verder. Dus ehh, ja het leidde over het algemeen tot een leefbaardere samenleving, ja, waarom? Ja, ik moet meer weten waarom. Waarom vinden ze dat? Dus daarom had ik hem een beetje, ja, het kan wel zo zijn, maar ik vond dan ook een beetje zwak (...) Omdat hij niet echt verder onderbouwd was.

I: En wat bedoel je daarmee, dat het... met onderbouwd?

R: Ehh, nou uitlegging waarom, uitleg waarom. Waarom ehh, waarom het überhaupt geeft voor ehh, zorgt voor een leefbaardere samenleving en ehh in wat, in wat voor opzicht dan.

Relevance criterion

A significant number of laymen (35%) doubt the relevance of the argument, so whether or not the argument is related to the matter at hand. Some focus on the applicability of the argument as a whole, as fragment (15) illustrates:

Fragment (15)

Evaluated argument (A):

Employees who have been ill for a long time and who get a holiday will generally return sooner to their regular workplace.

Supported claim:

If employers give employees who are absent due to long-term illness the opportunity to telework, they will return to their regular workplaces sooner.

In English:

R: And A, er, here I don't at all see the connection between these two reasonings, claims. Employees who have been ill for a long time and who get a holiday will generally return sooner to their regular workplace. That does not mean anything. Sooner than what? Employees who have been ill for a long time and who get a

holiday will generally return sooner to their regular workplace. Sooner than who or what, it doesn't say, and it also does not say as a result of what. So, as if it is a conclusion, if If employers give employees who are absent due to long-term illness the opportunity to telework, they will return to their regular workplaces sooner. Those two sentences have nothing to do with each other. So A comes last.

In Dutch:

R: En A, eh, daar vind ik de samenhang tussen deze twee stellingen, redeneringen, volstrekt verloren. Langdurig zieke werknemers die vrijaf krijgen zullen over het algemeen sneller terugkeren naar hun gewone werkplek. Dat zegt helemaal niks. Sneller dan wat? Langdurig zieke werknemers die vrijaf krijgen zullen over het algemeen sneller terugkeren naar hun gewone werkplek. Sneller dan wie of wat, dat staat er niet bij, en waardoor staat er ook niet bij. Dus, alsof het een conclusie is, als werkgevers langdurig zieke medewerkers in de gelegenheid stellen om te telewerken, dan zullen ze eerder terugkeren naar hun gewone werkplek. Die twee zinnen hebben niets met elkaar te maken. Daarom A als laatste.

---

Others focus in particular on the relation between the cause in the argument and the cause in the conclusion, as the person in fragment (16) does:

Fragment (16)
---------------

Evaluated argument (A):

Employees who have been ill for a long time and who get a holiday will generally return sooner to their regular workplace.

Supported claim:

If employers give employees who are absent due to long-term illness the opportunity to telework, they will return to their regular workplaces sooner.

In English:

R: Well, I put A here be, because that's the only place, you know, that was left over, so A is in third place because I do not, actually do not see anything in it. Er yeah employees who have been ill for a long time and who get a holiday will generally return sooner... yeah, I, a, an ill employee does not have a holiday, that er. I, I, you know, do not really see an argument in it. So er, I could not do anything with that.

In Dutch:

R: Nou, A staat hier om, omdat die enige plaats overbleef zeg maar, dus A staat op de derde plaats omdat ik er helemaal, eigenlijk helemaal niks in zie. Ehh ja langdurig zieke werknemers die vrijaf krijgen zullen over het algemeen sneller ... Ja, ik, een, een zieke werknemer heeft geen vrij, dat ehh. Ik, ik zie daar, ik zie er geen argument in, zeg maar. Dus ehh, daar, daar kon ik even niks mee.

---

The respondents in the current study never focused on the relevance of the effect only.

## (II) Specific to different argument schemes

### Both cause necessity and cause sufficiency criterion

Respondents who consider the strength of an argument from cause to effect may require both sufficiency and necessity of the cause (30%). Not only do they argue that if the cause occurs, then the effect will follow (sufficiency cue), but they also argue that without the cause occurring, the effect will not follow (necessity cue). For example:

#### Fragment (17)

Evaluated argument (E):

If ill employees can telework, then they will take on more work. Moreover, they can stay involved with work and colleagues. Eventually they will return to their regular workplace sooner.

Supported claim:

If employers give employees who are absent due to long-term illness the opportunity to telework, they will return to their regular workplaces sooner.

In English:

I: Well, I see you put E on 1. Can you explain why you put it there, why you thought E was the best?

R: Let me think for a while why that was, er [Thinking] Yeah, I think that if they do stay involved with work at home, that then they feel more like going back to work, eventually (...) If they still know how their colleagues are doing and stuff like that. Because if that contact fades away, then I don't think that they really feel like going back to work and on the contrary, they will sooner feel left out or something (...) And then they feel like working even less. So I think that, by doing telework, when they are ill, that that on the contrary er, that they will still feel like going to work.

In Dutch:

I: Nou, ik zie E op 1 staan. Kun je uitleggen waarom die daar staat, waarom je E het beste vond?

R: Even denken wat het ook alweer was, ehm [Denkt] Ja, ik denk dat als ze inderdaad thuis gewoon betrokken blijven bij het werk, dat ze dan nog meer zin hebben om weer terug te gaan naar het werk, op een gegeven moment (...). Als ze nog weten hoe het met hun collega's is enzo. Want als dat contact vervaagt, dan denk ik niet dat ze nog echt zin hebben om naar hun werk te gaan en dan zullen ze zich sneller juist buitengesloten voelen ofzo (...). En dan hebben ze er minder zin in. Dus ik denk dat, door telewerken, als ze eenmaal ziek zijn, dat dat juist wel dan ehh, dat ze dan wel nog zin hebben om naar hun werk te gaan.

The respondent in fragment (17) considers job involvement not only a sufficient cause, but also a necessary cause. It is argued that staying involved does indeed lead to a return to work (indicating sufficiency) and that not staying involved will not lead to a return to work (indicating necessity).

### Cause necessity criterion

A fifth of the laymen in this study use criteria referring to a necessary condition. For example:

#### Fragment (18)

Evaluated argument C:

Measures to let employees who have been ill for a long time work from home lead generally to the solution of social and individual problems.

Supported claim:

If employers give employees who are absent due to long-term illness the opportunity to telework, they will return to their regular workplaces sooner.

In English:

I: Okay, and then in fourth place you put C...

R: Yeah, let's see... Er, yeah I don't see it that way, that it, that working from home leads to the solution of social and individual problems. I think that it does not have so much to do with the workplace, whether it is at home or in the office. I don't really see the connection.

I: And why don't you really see the connection?

R: Er, well I think that social and individual problems are there and that they er can also be solved in all sorts of places, so that does not necessarily have to do with telework.

In Dutch:

I: Oké, en dan op de vierde plek heb je C staan...

R: Ja, even kijken... Ehm, ja ik zie dat niet zo, dat het, dat thuiswerken leidt tot het oplossen van maatschappelijke en individuele problemen. Ik denk dat het niet zozeer met de werkplek te maken heeft, of dat nou thuis is of op je werk. Het verband zie ik niet zo.

I: En waarom zie je dat verband niet zo?

R: Ehm, nou ik denk dat maatschappelijke en individuele problemen er zijn en dat die ook ehh op allerlei soorten plekken opgelost kunnen worden, dus dat heeft niet per se te maken met thuiswerken.

In fragment (18), the respondent makes clear that the proposed effect can occur without the proposed cause ever happening and that therefore, a causal link cannot be inferred.

### Effect desirability criterion

A few respondents (15%) focus on the desirability of an effect in the argument. In fragment (19), the respondent appreciates the argument because he or she considers the effect of involvement desirable:

#### Fragment (19)

Evaluated argument (E):

If more money is put at the disposal of youth drama schools, more young people will get the chance to become involved in society. Moreover, young people will

---

educate themselves better. Eventually, society will become more livable.

Supported claim:

Giving more money to youth drama schools will promote livability in society.

In English:

I: Okay, er, can you explain why you put E in 1<sup>st</sup> place?

R: Yes, er, I put E on 1, because it says, then more young people will get the chance to be involved in society. And I think that it is simply important that more young people are involved in society.

In Dutch:

I: Oké, ehm, kunt u uitleggen waarom E op 1 staat?

R: Ja, ehh, E heb ik op 1 gezet, omdat er staat, dan krijgen meer jongeren de kans om betrokken te worden bij de maatschappij. En ik denk dat het gewoon belangrijk is dat meer jongeren betrokken worden bij de maatschappij.

---

Table 3.11 shows the laymen criteria for the argument from cause to effect.

Table 3.11 Laymen criteria for the argument from cause to effect (per type of stimulus material) and agreement among respondents (the number of closed case individual interviews in which the criterion was mentioned divided by the total number of closed case individual interviews) (laymen criteria that do not correspond to theory are in italics).

Nature of criterion	Criterion	Type of stimulus material		Agreement among respondents, based on closed case individual interviews
		Open case	Closed case	
Specific to the argument from cause to effect	Cause sufficiency	+	+	19 of 20 (95%)
	Relevance (of the cause or the argument as a whole)	-	+	7 of 20 (35%)
Specific to different argument schemes	Both cause necessity and cause sufficiency	-	+	6 of 20 (30%)
	Cause necessity	-	+	4 of 20 (20%)
	<i>Effect desirability</i>	+	+	3 of 20 (15%)

### 3.4 Conclusion and discussion

#### Conclusion

In section 3.2, I examined the normative considerations on the quality of the argument from cause to effect in argumentation theory. In section 3.3, I uncovered the considerations of laymen on the quality of the argument from cause to effect. In this section, the key question is: to what extent do laymen norms as uncovered in the previous section correspond to those formulated in argumentation theory?

Almost all laymen in this study judge the argument from cause to effect on the basis of cause sufficiency and a smaller part of the laymen in this study judge this type of argument on the basis of cause relevance. Laymen's evaluations on the basis of cause sufficiency are generally on a more concrete level and less strict than suggested in theory. For instance, they do not ask the abstract question if a certain cause is sufficient to bring about the effect, but rather whether the specific effect in the argument (e.g. recovering and going back to work) is likely to occur as a result of the specific cause in the argument (e.g. telework). In some cases, laymen also use tests for cause sufficiency that are suggested in theory. To determine cause sufficiency, for example, they concentrate on the presence of contributing and/or counteracting factors and on the strength of the causal generalization.

In argumentation theory, evaluation criteria have been suggested for causal reasoning that can be considered irrelevant to the argument from cause to effect. Some laymen do indeed use these irrelevant criteria: some question whether a cause is both a necessary and sufficient condition, others question if the cause is a necessary condition. However, it should be noted that only a minority of the respondents use these irrelevant criteria.

In sum, the laymen criteria that are also mentioned in argumentation theory are the following:

Corresponding criterion with relatively high agreement among respondents:

- The cause sufficiency criterion: the cause should be enough to probably bring about the effect (specific to the argument from cause to effect).

Corresponding criteria with relatively low agreement among respondents:

- The cause relevance criterion: the cause in the argument should connect adequately to the cause in the conclusion (specific to the argument from cause to effect).

- The both cause necessity and cause sufficiency criterion: the cause should not only be enough to probably bring about the effect, but should also be a condition without which the effect does not occur (unexpected, because irrelevant to the argument from cause to effect).
- The cause necessity criterion: the cause should be a condition without which the effect does not occur (unexpected, because irrelevant to the argument from cause to effect).

There are also differences between argumentation theory and laymen. Most of these differences, however, could be expected. Laymen do not attack the strength of the conclusion, as suggested in theory. This does not come as a surprise, because in the material, the conclusion did not differ from argument to argument. So, if they would have had this particular criterion in mind, they were not stimulated to use it.

Other differences between theory and laymen could be expected, considering the irrelevance of the particular criteria to the argument from cause to effect. Respondents did not mention that association should not be necessarily interpreted as causality, that other effects than those under consideration should not be (un)desirable or that there should not be better actions than the one proposed.

The cause accuracy criterion was not used by laymen, as expected. This criterion only applies to existing causes, not to causes that are hypothetical. In the material, only hypothetical causes were presented, so respondents did not have any reason to doubt them.

Finally, laymen in this study base their judgments upon the desirability of an effect in the argument, something that was not suggested in theory for the argument from cause to effect and that can also be considered irrelevant to this type of argument. However, only a small part of the respondents used this criterion. In sum, the non-corresponding criteria are:

Non-corresponding criteria that are not mentioned by respondents:

- The strength of conclusion criterion: the conclusion should not be too strong (in theory only, as expected, because of the material).
- The association or causation criterion: association should not be necessarily interpreted as causality (in theory only, as expected, because irrelevant to the argument from cause to effect)
- The nature of side-effects criterion: other effects than those under consideration should not be (un)desirable (in theory only, as expected, because irrelevant to the argument from cause to effect).
- The counterplan criterion: there should not be better actions than the one proposed (in theory only, as expected, because irrelevant to the argument from cause to effect).



- The cause accuracy criterion: the cause should not be doubted (in theory only, as expected, because irrelevant to the hypothetical cause).

Non-corresponding criterion with relatively low agreement among respondents:

- The effect desirability criterion: the effect in the argument should be desirable (laymen only, unexpected, because irrelevant to the argument from cause to effect).

## Discussion

In general, laymen use norms that they are supposed to use according to those criteria formulated in argumentation theory, although the cause sufficiency criterion is the only criterion shared by a majority of laymen. Some norms that laymen use for the argument from cause to effect are in fact irrelevant to this type of argument. Some focus on the cause as a necessary condition and others on the desirability of the effect in the argument, something that they should not do for the argument from cause to effect. Rather, these criteria pertain to the argument from effect to cause and pragmatic argumentation, respectively.

The finding that laymen appear to use irrelevant criteria should be qualified: the agreement on these criteria was relatively low and it is understandable that they were used. The tendency to focus on the desirability of an effect in the argument can easily be explained by the pragmatic context created in the material. In the material, each of the presented claims was an action-consequence relation that could be used as a premise in a pragmatic argument. If someone considers an action and its consequence in the argument, it may be difficult to ignore the desirability of the effect, even though it is not relevant to the reasonableness of the argument. Similarly, the focus on the necessity of the cause might be inherent in the pragmatic context; if a certain (desirable) effect is presented, then people might be stimulated to think about alternative actions that might also lead to the effect.

The results of the closed case interviews showed that some laymen apply the notion of sufficient conditions more flexibly than they should according to formal logicians and some argumentation theorists. This result is in line with research by Verschueren, Schroyens, Schaeken and d'Ydewalle (2004). They investigated how lay people understand and assess the notions of necessity and sufficiency of a cause for the effect. Specifically, it was experimentally determined if the subjective conceptualization of necessity and

sufficiency paralleled the formal concept. For sufficiency, the formal indications were (a) when the cause occurs the effect follows and (b) the absence of the effect implies the absence of the cause. For necessity, the indications were (c) the occurrence of the effect implies the occurrence of the cause and (d) the absence of the cause implies the absence of the effect.

Verschueren et al. (2004) reported that the laypeople's notions are not completely identical to the formal definitions: they interpreted sufficiency in a more liberal (or graded) way whereas necessity was considered as an all-or-none affair. They concluded that it "seems to be more ecologically valid to define sufficiency in terms of for instance the likelihood that the cause brings about the effect, rather than in terms of an all-or-none property" (p. 13). After all, so they argue, in everyday language it is rarely the case that the cause is always followed by the effect, as in strict sufficiency<sup>67</sup>. This conclusion corresponds to the results of the study reported in this chapter, as the closed case interviews revealed that laymen may think in terms of probability when they judge causation and not necessarily in strict sufficiency. It also corresponds to the probabilistic approach to causality, in which it is assumed that events might change the probability of other events (see 3.1.1).

If the laymen criteria are compared with the approaches to causality that Hulswit (2002) discussed (see 3.1.1), it can be concluded that all different approaches are represented in the laymen criteria. In the results of this study, all contemporary philosophical accounts of causality show through: (I) a necessary and/or sufficient conditions approach, (II) a counterfactual approach, (III) an instrumental or agency approach and (IV) a probabilistic approach.

A limitation of the study reported in this chapter is the design of the open cases. The data generated by the open case were not as informative as expected. The pragmatic context might have complicated the task. In creating arguments and evaluating them, people were somehow stimulated to focus on the proposed measures (reducing fat content in minced meat and improving the greenery in new housing estates) instead of the prediction that the proposed measure will lead to a consequence. So, they were tempted to

---

<sup>67</sup> This point of view can be found in argumentation theory. Hastings (1962) already argued: "The logician would have us require sufficient cause to obtain a certain relation, but in rhetoric we can usually only approach sufficient cause, since we deal with relations of probability" (p. 76). Schellens and Verhoeven (1994) must have had the same feeling, considering the critical question: "Is cause A in general sufficient to consider effect B probable?" (p. 82).

present an argument in favor of the action in the claim and an argument against the action in the claim. The instruction and interview techniques might have been insufficient to prevent or correct that problem (the respondents were not corrected during the interview).

It was observed that respondents in this study never focused on the relevance of the effect only. However, this can be explained by overly subtle manipulations of the effects in the closed case. More specifically: the solution of social and individual problems was insufficiently unrelated to employees returning to their regular workplaces and an enduring future was insufficiently unrelated to a promoted livability of society. In that respect, the manipulations in the closed case might have been inadequate to reveal that laymen focus on the relevance of the effect in the argument.

Still, the results of this study are valuable, as insights are gained into the specific features that should make the argument from cause to effect strong. The ideas about what should make this argument type strong do not differ considerably between argumentation theory and laymen, although it should be noted that the sufficiency criterion is the only criterion that laymen in this study agree upon to a relatively high extent.

## 4. The quality of the argument from example

This chapter deals with the quality of the argument from example. First, in section 4.1, I will discuss various functions of the example and indicate which function of the example is relevant to this study. Then, I will address the questions of what criteria are used in argument theory to evaluate the quality of the argument from example (4.2) and what criteria are used by laymen to evaluate the quality of the argument from example (4.3). Finally, I will make a comparison between these conceptions (4.4).

### 4.1 Introduction

#### 4.1.1 Functions of an example

Just like authority, 'example' may mean different things in different situations. Willer, Ruchatz and Pethes (2007) provide a systematic treatise on the functions of an example that abstracts from different research traditions. They present a typology of functions of the example. In categorizing these functions, they focus on the nature of the example as a source of knowledge ('the epistemology of the exemplary'). I will review the typology developed by Willer et al. (2007) and use it as a starting point to narrow down my treatment to the type that is relevant to this study. In Figure 4.1-4.4, I present the diagrams Willer et al. (2007) use to represent the different functions of the example. The diagrams have been copied directly from the originals, but the captions have been translated from German into English.

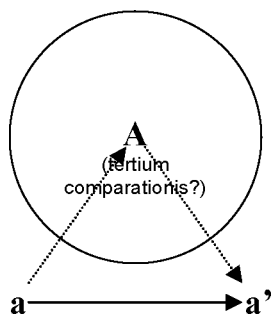


Figure 4.1 The rhetorical example

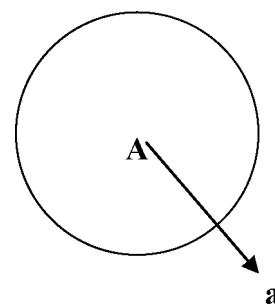


Figure 4.2 The knowledge-representing example

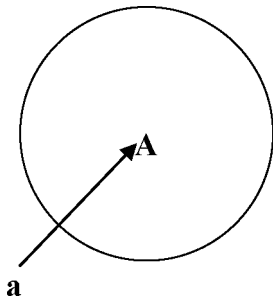


Figure 4.3 The knowledge-constructing example

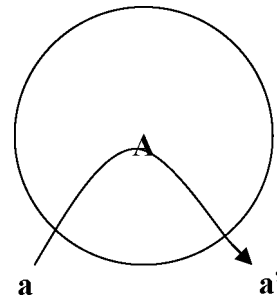


Figure 4.4 The normative example

The first function of an example distinguished by Willer et al. (2007) is *the rhetorical function* (“rhetorische Funktion” / “das rhetorische Beispiel”), referring to the function of the example as it has been treated in the rhetorical tradition (see Figure 4.1). In the traditional rhetorical treatment, the example is a particular case *a* (e.g. a story, a person, a historical event) from which a conclusion is drawn for a separate particular case *a'*. The reasoning may go directly from one particular case to another particular case, but may also proceed (explicitly or implicitly) via a general rule *A* or a *tertium comparationis*, the quality or qualities that the compared particulars have in common. To clarify the rhetorical function further and to show the function in an argumentative context, I present the following examples<sup>68</sup>:

- (1) *a*: The other library I went to last week offered free web access.  
*(A: All libraries offer free web access. /*  
*tertium comparationis*: The library I'm about to visit and the one I went to last week have relevant qualities in common.)  
*a'*: Therefore, the library I'm about to visit will probably offer free public internet access too.
  
- (2) *a*: The library I went to last week was right to offer free web access.  
*(A: All libraries are right to offer free web access. /*  
*tertium comparationis*: The library I'm about to visit and the one I went to last week have relevant qualities in common.)  
*a'*: Therefore, this library would be right to offer free public internet access too.

<sup>68</sup> All examples in (4.1.1) are created by myself and not by Willer et al. (2007).

On the basis of Willer et al. (2007), I could not determine the nature of the rule *A* that the rhetorical example may rely on, but the examples above show that the rule can be either descriptive (example 1) or normative (example 2). The examples also show that Willer et al.'s (2007) rhetorical example is in fact an argument by analogy in an argumentative context.

The second function Willer et al. (2007) ascribe to the example is *the knowledge-representing function* ("wissenabbildende Funktion" / "Das Belegbeispiel", see Figure 4.2). In this function, example *a* serves as an illustration of a known rule *A*, a rule to which it at the same time contributes. In other words, an element of a group is used to make general regularities of this group concrete. The knowledge-representing example appeals to descriptive rules only, expressing what *is* (see also Willer et al., 2007, p. 42). In my opinion, the knowledge-representing function might play a role in an argumentative context, but the question is whether it is an argumentative role. I will return to this point at the end of this section.

Willer et al. (2007) call the third function of the example *the knowledge-constructing function* ("wissenbildende Funktion" / "Das Ausgangsbeispiel", see Figure 4.3). The authors explain that the knowledge-constructing example is the mirror image of the knowledge-representing example. In the former, we start from a more general knowledge and illustrate this knowledge with a more specific example, whereas in the latter the more specific knowledge (the example) is used to construct the more general knowledge. For example:

- (3)     *a*: Yesterday you received unwanted messages and today you clicked on an attached Trojan horse.  
          *A*: Therefore, most emails are spam.

According to Willer et al. (2007, p. 42), the knowledge-constructing example makes reference to descriptive rules only, just like the knowledge-representing example. The example shows that the knowledge-constructing example becomes reasoning from example to a generalization in an argumentative context.

*The normative function* of the example is the last function identified by Willer et al. (2007) ("normative Funktion" / "Das normative Beispiel", see Figure 4.4). In this role, something or somebody serves as a model to be imitated - or not be imitated in case the model is bad. Things that may serve as normative examples range from artworks to narratives; human normative examples may vary from teachers to saints. Willer et al. (2007) explain that the

normative example is used to present abstract norms in a clear and attractive way to make them applicable to other individuals. The next example shows the normative function in an argumentative context:

- (4) A: Managers should attempt to influence motivation.  
a: Consider our former manager: he was recognized by his employees as an exceptional leader.  
a': Therefore, as our new manager, you should try to be a motivator too.

Example 4 shows that the normative example only applies to prescriptive rules, rules about what *ought* to be (see also Willer et al., 2007, p. 42). The normative example appears to be strongly related to the knowledge-representing example, because something general ('Managers should attempt to influence motivation') is illustrated by a specific example ('our former manager'). In my opinion, one difference between the knowledge-representing example and the normative example is that the former illustrates a descriptive rule, whereas the latter illustrates a prescriptive rule. Another difference is that the knowledge-representing example is used with the purpose to illustrate, whereas the normative example is used with the ultimate goal to draw a conclusion for another particular case, which makes it similar to the rhetorical example. A difference, however, is that the rhetorical example does not have an illustrative function, whereas the normative example does. After all, the normative example is used to illustrate a prescriptive rule. About the rhetorical example, it was stated that it is in fact an argument by analogy in an argumentative context. The normative example as presented by Willer et al. (2007) on the other hand is in fact *not* an argument by analogy in an argumentative context, as it is not reasoned from a particular case to another particular case. In other words, the normative example in an argumentative context does not form a premise in the argumentation. Rather, it is reasoned from a prescriptive rule (illustrated by the normative example) to another particular case.

What function of the example is relevant within the framework of pragmatic argumentation? In chapter 1, I explained that I would focus on the evaluation of argument types supporting one of the premises in the pragmatic argumentation scheme: the premise expressing that action A (probably) leads to B. This premise can be conceived as a general rule - or rather a *causal generalization* - and it is possible to support such a causal generalization with one or more examples exhibiting the causal relation. For instance: "In Rotterdam, the introduction of a public transport chip card has led to better public

transport. (In Amsterdam etc., the introduction of a public transport chip card has led to better public transport.). Therefore, introducing the public transport chip card usually improves public transport.”. Now I can specify the question further into: what function of the example is relevant to the causal generalization?

If we consider the functions distinguished by Willer et al. (2007), it appears that the rhetorical example and the normative example (Figure 4.1 and 4.4) are not relevant to the causal generalization, as they are used to support a claim about a particular case instead of a general principle. The knowledge-constructing and knowledge-representing example (Figure 4.2 and 4.3) appear to be more relevant: in the former, reasoning goes from one or several examples to a general principle whereas in the latter, a general principle is illustrated by one or several examples. The example can do both jobs for the causal generalization in the pragmatic argumentation scheme: it may justify that action A (probably) leads to B or it may illustrate that action A (probably) leads to B.

Although the knowledge-representing function appears to be relevant to the causal generalization in pragmatic argumentation, it is not relevant to the current research question. Considering this research question, I am interested in the quality of *argumentation*. By using the knowledge-representing example, however, someone does not aim to increase the acceptability of a controversial standpoint for the listener or reader, nor is the example put forward to justify the standpoint<sup>69</sup>. By using the knowledge-representing example, someone aims to *illustrate*<sup>70</sup>, something I am not interested in within the framework of this study.

In addition, argumentation theory gives us insufficient reason to think that an illustrative example is in fact argumentative. According to Kienpointer (1992), the illustrative example can be considered to be argumentation, as appears from the difference that he makes between *Induktive Beispielargumentation* and *Illustrative Beispielargumentation*. In the first type of argument, a general rule

---

<sup>69</sup> I rely on the following definition of argumentation: “*Argumentation is a verbal and social activity of reason aimed at increasing (or decreasing) the acceptability of a controversial standpoint for the listener or reader, by putting forward a constellation of propositions intended to justify (or refute) the standpoint before a rational judge*” (Italics in original, Van Eemeren, Grootendorst & Snoeck Henkemans, 1996, p. 5).

<sup>70</sup> According to the Oxford English Dictionary (online second edition, 1989), illustration can be defined as “The action or fact of making clear or evident to the mind; setting forth clearly or pictorially; elucidation; explanation; exemplification” (retrieved online, March 4<sup>th</sup>, 2009).



that has not been established yet is induced from one or more examples. Therefore, this is similar to the knowledge-constructing function of an example. In the second type, one or more examples are used to illustrate or to confirm a general rule serving as a premise in an argumentation. This is similar to the normative example, which is not an argument, but an illustration of a rule that is an argument. I disagree with Kienpointer's (1992) suggestion that an illustrative example is argumentative, because it does not form a premise upon which the conclusion is based. Instead, the illustrated rule is the argument here.

Schellens (1985) suggests that an illustrative example is not an argument: "In some cases it is the question if the presumption of argumentation is rightfully ascribed to examples. Does the writer give the example to support a general statement, or does he just want to illustrate the statement?"<sup>71</sup> (p. 192, translated from Dutch). Walton (1996) appears to agree with Schellens (1985): "One must be careful in identifying argument from example in a given case, because the citing of an example is also used in explanations and other speech acts that are not arguments" (p. 51).

Although it is not always possible in concrete cases to determine if an example has an illustrative or an argumentative function, only the knowledge-constructing example is in my opinion relevant to the causal generalization, as the specific case is used as support for the knowledge that action A (probably) leads to B. Therefore, I will limit my discussion to the knowledge-constructing example. Next, I will consider the quality of the knowledge-constructing example according to common views in the scientific community.

#### **4.1.2 Sampling principles in scientific methodology**

Inducing general principles from particular examples is linked to scientific, inductive ways of reasoning. In this scientific approach, the particular example is usually part of a sample. Sampling may be part of quantitative research, aiming to make (statistical) inferences about a population. The requirements for sound quantitative sampling have become a matter of common knowledge in modern science. The sample ought to be representative of the population, which means

---

<sup>71</sup> In Dutch: "In sommige gevallen is het de vraag of aan voorbeelden met recht de pretentie van argumentatie wordt toegeschreven. Geeft de schrijver het voorbeeld ter verdediging van een algemene uitspraak, of wil hij er de uitspraak alleen maar mee illustreren?" (1985, p. 192).

that it should give an accurate representation of the population's distribution. The representativeness depends on two factors: the size of the sample and the random selection of the members of the sample. The larger the sample, the higher the chance that it truly reflects the population. A sample is chosen randomly, if every member of the population is given an equal and known chance of being selected for the sample.

Sampling strategies are not only used in quantitative research methods, but also in qualitative research methods. As qualitative samples are not meant to be statistically representative, different sampling principles have been developed. According to Ritchie, Lewis and Elam (2003), the first principle of qualitative sampling is the requirement for 'symbolic representation', as opposed to statistical representation required in quantitative research. According to this sampling principle, "Units are chosen because they typify a circumstance or hold a characteristic that is expected or known to have salience to the subject matter under study" (pp. 82-83). This corresponds with what Shadish, Cook and Cambell (2002) call 'purposive sampling of typical instances': "the aim is to explicate the kinds of units, treatments, observations, and settings to which one most wants to generalize and then to select at least one instance of each class that is impressionistically similar to the class mode" (p. 24).

The second principle listed by Ritchie et al. (2003) is the requirement for diversity. They explain why a qualitative sample needs to be diverse: "it optimises the chances of identifying the full range of factors or features that are associated with a phenomenon [and] it allows some investigation of interdependency between variables such that those that are most relevant can be disengaged from those of lesser import" (p. 83). Shadish et al. (2002) refer to the same principle discussing the 'purposive sampling of heterogeneous instances': "the aim is to include instances chosen deliberately to reflect diversity on presumptively important dimensions, even though the sample is not formally random" (p. 23).

Shadish et al. (2002) explain and illustrate the difference between the two sampling options as follows:

"The starting point for purposive sampling of heterogeneous instances is the same as the starting point for purposive sampling of typical instances – defining the characteristics of the persons, settings, treatments, or outcomes to which one wants to generalize. However, whereas PSI-Typ [purposive sampling of typical instances] aims to create a sample that *is* typical on those characteristics, PSI-Het [purposive sampling of heterogeneous instances] aims to create a sample that is heterogeneous

and that need not include a typical instance at all. For example (...), if one's goal was to study elderly residents of nursing homes, PSI-Type would lead the researcher to include residents whose ages were all about, say, the mean age of nursing home residents in the population of interest, or at least whose average age was about the mean of all residents. But PSI-Het would lead the researcher to select patients whose ages vary widely, from the youngest to the oldest such patients. Not only would PSI-Het *not* require including a patient whose age was actually at the mean, it would not even require the average age of the sample to match that mean" (p. 376).

Translated to the argument from example: if we want to defend the claim that elderly residents of nursing homes show health improvements if they are visited by their children more often, we may choose typical instances and present examples of residents whose ages are close to the mean age of all residents, e.g. residents aged 77, 75 and 73. The other option would be to select heterogeneous examples, e.g. residents aged 65, 72 and 80.

In sum: according to quantitative sampling principles, a sample formed by particular examples should be statistically representative. According to qualitative sampling principles, it is considered desirable that sampling units stand for the object under investigation (typicality) or are as diverse as possible (diversity).

#### 4.2 Argumentation-theoretical criteria

In argumentation theory, the argument from example is often treated as inductive generalization. The frequent association between the argument from example and inductive generalization has been observed by, for instance, Schellens (1985), who states that "All treatments, more or less directly, also make a connection with induction on the basis of a sample"<sup>72</sup> (p. 190). In more recent years, Arthos (2003) remarked along similar lines: "Virtually all of the available textbooks on argumentation written from a rhetorical perspective teach argument from example as a species of inductive generalization" (p. 320). It has been argued that this conception of the argument from example is problematic, because in persuasive texts inductive generalizations usually occur in the form of one or more typical examples (Schellens, 1985).

In my opinion, treating the argument from example as inductive generalization is not necessarily problematic, as long as it is not treated as the statistical kind of inductive generalization

---

<sup>72</sup> Translated from Dutch: "Alle behandelingen leggen ook meer of minder direct een verband met inductie op basis van een steekproef" (1985, p. 190).

connected to statistical representativeness. If the argument from example is treated as inductive generalization of the qualitative kind - arriving at general or universal propositions on the basis of a purposive sample - there should not be a problem. After all, in purposive sampling, the instances are selected subjectively, because they appear typical or heterogeneous (see also 4.1.2). In that respect, it resembles the form that the argument from example takes in persuasive texts.

In argumentation theory, there is not much evidence that the argument from example is approached as a statistical kind of inductive generalization. Here is a selection from the various characterizations: "The process of reasoning by example consists of inferring conclusions from specific cases (...) Sometimes a single case may be used to establish the conclusion or generalization. More often a number of cases will be offered as the basis for the conclusion. Reasoning by example is a form of inductive reasoning (...)" (Freeley & Steinberg, 2000, pp. 153-154), "In a generalization one reasons that what is true of certain members of a class will also be true of other members of the same class or of the class as a whole (...) The generalization (...) involves more than two instances and often makes claims about a whole class of objects" (Warnick & Inch, 1989, p. 105)", "In this process one or more examples are examined and then a generalization is made describing all examples of the same type. The conclusion is not just a description of the examined examples, but applies to examples not observed" (Windes & Hastings, 1969, p. 168). From the various definitions, it appears that the authors do not necessarily refer to inductive generalization of the statistical kind<sup>73</sup>.

In argumentation theory, critical questions have been formulated to evaluate the quality of the argument from example. In general, no distinction is made between reasoning from example to a

---

<sup>73</sup> It can be argued that these authors refer to inductive generalization of the statistical kind in their evaluation questions: they often require representativeness and a sufficient number of examples (see also section 4.2.2). However, it is uncertain whether they mean *statistical* representativeness and whether they require a sufficient number of examples, because it increases statistical representativeness. They might as well require representativeness of the *qualitative* kind, and a sufficient number of examples to create diversity in the argument. In addition, if they would aim at inductive generalization of the statistical kind, they would probably also mention the criterion of *random selection*, as that criterion is - like sample size - associated with statistical representation. Therefore, I maintain the conclusion that they do not necessarily refer to the statistical kind of inductive generalization.

*descriptive generalization* and reasoning from example to a *causal generalization* (except for Windes & Hastings, 1969, pp. 168-173). Instead, the two kinds are grouped together in argumentation theory and 'disappear' under the label *argument from example (to a generalization)*.

In this section, the key question is: what evaluation questions have been suggested in argumentation theory for the evaluation of the argument from example? To address this question, a literature study was carried out.

#### **4.2.1 Method**

I refer to section 3.2.1 for a description and justification of the selection of literature. The same literature was employed in this study.

#### **4.2.2 Results**

In this section, I will report on the criteria that have been formulated in argumentation theory for the argument from example. One group of criteria is specific to the argument from example (I) and another group of criteria not only applies to the argument from example, but also to other argument schemes (II).

##### (I) Specific to the argument from example

##### (I.1) Typicality/representativeness criterion

The words 'typical' and 'representative' occur frequently in the evaluation questions (see Table 4.1). Often, however, it is not clear how these terms should be understood. Should the examples, for instance, give an accurate representation of the population distribution, or should they embody the essential characteristics of the matter under discussion? The elaborations of the authors are not always clear in this respect. Freeley and Steinberg (2000), for instance, explain the question "Are the examples typical?" as follows: "The advocate must determine whether the cases offered are really representative" (p. 155). Garssen (1997) claims that to answer the question whether or not the examples are typical, "the antagonist should first wonder whether the individual cases are representative and whether enough individual cases are considered"<sup>74</sup> (pp. 11-12,

---

<sup>74</sup> In Dutch: "(...) zal de antagonist zich eerst moeten afvragen of de afzonderlijke gevallen wel representatief zijn en of er genoeg afzonderlijke gevallen in ogenschouw zijn genomen" (1997, pp. 11-12).

translated from Dutch). He does not explain the terms 'typical' and 'representative' any further and does not clarify why the antagonist should ask the preliminary questions before being able to judge the typicality of the examples.

Schellens and Verhoeven (1994), on the other hand, define 'typical': "(...) examples can be *typical* for a class for various reasons: an example can be part of a representative sample; it can be a representative of the average in a class and it can be an extreme example that takes a known position in relation to the rest of the class"<sup>75</sup> (p. 136, translated from Dutch). To illustrate that an example can be typical thanks to its extreme position, Schellens and Verhoeven (1994) use the following example: "The last chemistry exam for pre-university education was certainly not difficult. Even Joris got a 7"<sup>76</sup> (p. 135, translated from Dutch). The word 'even' in this example indicates that Joris usually gets marks lower than 7 and that his mark for the last chemistry exam is unusual, or extreme. So it can be reasoned that if the exam was not even difficult for Joris, it certainly was not difficult for the other examinees. Thanks to its extreme position, the case of Joris is typical for the exam's difficulty level.

Apart from the problem that the terms 'typical' and 'representative' are generally loosely defined in argumentation theory, it can be challenged – and it has been challenged - that examples in the argumentation should reflect the population well (e.g. Schellens, 1985, pp. 193-194). One single example, or two, or three, will not succeed in giving such an accurate representation, unless the population is homogeneous.

---

<sup>75</sup> In Dutch: "(...) voorbeelden om verschillende redenen *typerend* kunnen zijn voor een klasse: een voorbeeld kan deel uitmaken van een representatieve steekproef; het kan een representant zijn van het gemiddelde in een klasse en het kan een extreem voorbeeld zijn dat een bekende positie inneemt ten opzichte van de rest van de klasse" (1994, p. 136).

<sup>76</sup> In Dutch: "Het laatste examen scheikunde voor het VWO was beslist niet moeilijk. Zelfs Joris had er een 7 voor" (1994, p. 135). For a correct interpretation of this example, it is relevant to mention that in the Netherlands, grades from 1 up to 10 are generally used, with 1 being the lowest and 10 being the highest.

Table 4.1 Evaluation questions falling under the category of (I.1) Typicality/representativeness criterion.

<i>Evaluation question</i>	<i>Source</i>
I.1 Typicality/representativeness criterion	
Are the examples typical for the class the conclusion relates to? <sup>77</sup>	Schellens (1985, p. 196); Schellens & Verhoeven (1994, p. 136)
Are the examples typical or representative of the class of objects about which the generalization is made?	Warnick & Inch (1989, p. 107)
Are the examples representative?	Reinard (1991, p. 196)
Is the example typical of the kinds of cases that the generalization ranges over?	Walton (1996, p. 50)
Are the specific cases typical for the state of affairs? <sup>78</sup>	Garssen (1997, p. 11)
Are the individual cases representative? <sup>79</sup>	Garssen (1997, p. 12)
Are the examples typical?	Freeley & Steinberg (2000, p. 155)
Are the examples typical of the category the speaker wants to generalize about?	Meany & Shuster (2002, p. 61)
I.1.1 Special circumstances criterion	
Is there different information available that makes the conclusion less probable? <sup>80</sup>	Schellens & Verhoeven (1994, p. 136)
Were there special circumstances present in the example that would impair its generalizability?	Walton (1996, p. 50)
I.1.1.1 Critical period criterion	
Do the examples cover a critical period of time?	Reinard (1991, p. 196)
Do the examples cover a critical period of time?	Freeley & Steinberg (2000, p. 155)

#### (I.1.1) Special circumstances criterion

Walton (1996, p. 50) proposes the question: "Were there special circumstances present in the example that would impair its generalizability?" (see Table 4.1). He adds: "Presumptive

<sup>77</sup> In Dutch: "Zijn de voorbeelden typerend voor de klasse waarop de conclusie betrekking heeft?" (1985, p. 196; 1994, p. 136).

<sup>78</sup> In Dutch: "Zijn de specifieke gevallen typerend voor de betreffende stand van zaken?" (1997, p. 11).

<sup>79</sup> In Dutch: "Zijn de afzonderlijke gevallen representatief?" (1997, p. 12).

<sup>80</sup> Not mentioned in Schellens (1985), only in Schellens and Verhoeven (1994). In Dutch: "Is er andersoortige informatie beschikbaar die de conclusie minder aannemelijk maakt?" (1994, p. 136).

generalizations are inherently defeasible, or subject to exceptions. One can always examine a cited example to look for indications that special circumstances are present in it” (p. 50). It can be argued that this question is a heuristic one, helping someone decide whether or not an example is typical of the kinds of cases the generalization ranges over. After all, from the fact that a case has particular circumstances it follows automatically that the example is atypical – unlike other instances of the same class.

Schellens and Verhoeven (1994) suggest the question “Is there different information available that makes the conclusion less probable?” (p. 136, translated from Dutch). If we interpret ‘different information’ as ‘other information than presented in the examples’, they refer, among other things, to the presence of special circumstances in the example. It is unlikely that they also refer to counterexamples, as they mention counterexamples in a different question. So, the question appears to be similar to Walton’s (1996) question “Were there special circumstances present in the example that would impair its generalizability?” (p. 50). Again, this test can serve to answer the question of typicality.

#### (I.1.1.1) Critical period criterion

As Table 4.1 shows, the question “Do the examples cover a critical period of time?” was proposed by Freeley and Steinberg (2000, p. 155), and by Reinard (1991, p. 196), who cites Freeley. Freeley and Steinberg (2000) explain why this question is important: “In many cases the time at which the examples were studied or the time period covered by the examples may be critical. The advocate should try to find examples representative of the period of time critical to the argument” (p. 155). This question appears to be a way to answer the question whether or not the example is typical or representative. If the example does not cover a critical period of time, it is not.

If the critical period question is answered with a ‘no’ (‘no, the examples do not cover a critical time period’), it means that there are special circumstances present in the example impairing its generalizability. After all, the particular time period covered by the examples form a special condition that cannot be generalized to the critical time period. Therefore, in the classification of criteria, the critical period category is a subcategory under the special circumstances category.



### (I.2) Number of examples criterion

It has been suggested in argumentation theory that enough examples should be given to support a claim (see Table 4.2). Warnick and Inch (1989), for instance, explain that the number of examples that should be given “depends on the nature of the audience and the situation”, but “in many cases the greater the number of examples the more cogent and persuasive the generalization will become” (p. 107). According to Freeley and Steinberg (2000), “advocates have to present enough cases to convince a reasonable person that there is a high degree of probability that a conclusion is correct” (p. 154). Garssen (1997) suggests that the question whether or not enough cases are cited should *precede* the question of typicality. Why this is necessary, is not clear. It seems to me that the typicality of the given examples can be judged regardless of the total number of examples.

Others, however, deny that giving enough examples is a fair requirement for the argument from example (e.g. Schellens, 1985; Walton, 1996). They note that it is typical for this type of argumentation that only a few examples are given and that the number of examples should not be a reason to reject the argument.

If the criterion of sufficient examples has been suggested to increase statistical representation - the phrasing of the critical questions and clarifications does not give a definite answer - it is remarkable that the criterion of random selection, indicating that each element should have an equal chance of being chosen, has not been proposed. After all, in sampling theory, statistical representation depends on both sample size and random selection. Windes and Hastings (1969) did mention the method of random selection (pp. 168-173). They also pointed to the diversity of characteristics in the examples, the principle common to qualitative sampling (see 4.1.2). It is striking that this criterion is not mentioned in more recent literature.

Table 4.2 Evaluation questions falling under the category of (I.2) Number of examples criterion.

<i>Evaluation question</i>	<i>Source</i>
I.2 Number of examples criterion	
Are there enough examples to reliably predict a general trend or characteristic?	Warnick & Inch (1989, p. 107)
Are enough examples cited?	Reinard (1991, p. 196)
Are there enough individual cases mentioned? <sup>81</sup>	Garssen (1997, p. 12)
Is there a reasonable number of examples?	Freeley &

<sup>81</sup> In Dutch: “Worden er genoeg afzonderlijke gevallen genoemd?” (1997, p. 12).

	Steinberg (2000, p. 154)
Are there enough examples to prove the claim?	Meany & Shuster (2002, p. 61)

### (I.3) Counterexamples criterion

The next test that has been suggested is to ask whether there are any examples that undermine the generalization (see Table 4.3). According to Warnick and Inch (1989), "In some cases, even one powerful counterexample can undermine the most carefully constructed generalization" (p. 107). Schellens (1985) explains that the effect of a counterexample depends on the form of the generalization. The more universal the conclusion, the more damage a counterexample can inflict. The conclusion that *all* party guests were amused can be falsified by giving a single counterexample, whereas such a counterexample is insufficient to refute the conclusion that *most* party guests were amused.

Walton (1996) and Garssen (1997) do not include the counterexample-test in their lists of critical responses. Nevertheless, Walton (1996) notes elsewhere in his discussion of the argument from example that "One type of critical response to [the] use of the argument from example is to present a counterexample - another example that falsifies or refutes the given generalization" and that "the response to this reply is for the proponent of the original generalization to qualify it with terms like "usually," "sometimes," "rarely," and so forth" (p. 50).

Table 4.3 Evaluation questions falling under the category of (I.3) Counterexamples criterion.

<i>Evaluation question</i>	<i>Source</i>
I.3 Counterexamples criterion	
Are there any counterexamples that undermine the conclusion in the given form? <sup>82</sup>	Schellens (1985, p. 196); Schellens & Verhoeven (1994, p. 136)
Are examples that counter the claim made in the generalization noncritical?	Warnick & Inch (1989, p. 107)
Are contrary examples unimportant?	Reinard (1991, p. 196)
Are negative examples noncritical?	Freeley & Steinberg (2000, p. 155)
Are there examples that might directly counter the given	Meany & Shuster

<sup>82</sup> In Dutch: "Zijn er tegenvoorbeelden die de conclusie in haar gegeven vorm ondergraven?" (1985, p. 196; 1994, p. 136).

examples?	(2002, p. 61)
-----------	---------------

#### (I.4) Strength of conclusion criterion

Walton (1996) suggests the question “How strong is the generalization?” (p. 50, see also Table 4.4). He indicates how this should be understood: “If the generalization holds only for a narrow range of cases, it is not very reliable. In such cases, the argument may hold, but may be weak” (p. 50). A comparable critical question comes from Schellens and Verhoeven (1994): “Are the certainty and the range of the conclusion not larger than made probable by the examples?” (p. 136, translated from Dutch). The questions appear to be similar, because they both refer to the strength of the conclusion. Even if there is nothing wrong with the given example(s), the conclusion may be formulated in such absolute terms that it is not acceptable. For instance: for the generalization ‘At all times, all students work harder for cash’, the cases of Rose, Sophia and Carl are presented. If you know Rose, Sophia and Carl as typical students who always work harder if they receive cash rewards and you cannot think of one single counterexample – a student who is not at all times motivated by cash incentives – you still might find the generalization too strong, because you are not convinced that the characteristic holds invariably true for all students. So, the generalization in the conclusion can be formulated too strongly, even though the argument holds.

Table 4.4 Evaluation questions falling under the category of (I.4) Strength of conclusion criterion and (I.5) Example accuracy criterion.

<i>Evaluation question</i>	<i>Source</i>
<b>I.4 Strength of conclusion criterion</b>	
How strong is the generalization?	Walton (1996, p. 50)
Are the certainty and the range of the conclusion not larger than made probable by the examples? <sup>83</sup>	Schellens (1985, p. 196); Schellens & Verhoeven (1994, p. 136)
<b>I.5 Example accuracy criterion</b>	
Are the examples probable? <sup>84</sup> ; Are there any reasons to doubt the correctness of the examples? <sup>85</sup>	Schellens (1985, p. 196); Schellens & Verhoeven (1994, p. 136)

<sup>83</sup> In Dutch: “Zijn de stelligheid en de reikwijdte van de conclusie niet groter dan de voorbeelden aannemelijk maken?” (1985, p. 196; 1994, p. 136)

<sup>84</sup> In Dutch: “Zijn de voorbeelden aannemelijk?” (1985, p. 196).

Is the proposition presented by the example in fact true?	Walton (1996, p. 50)
---	----------------------

(I.5) Example accuracy criterion

Questions have been suggested regarding the accuracy of the example (see Table 4.4). These questions ask “whether the example is a true report of reality” (Walton, 1996, p. 50). For the argument from cause to effect, a similar question was formulated: Are there any reasons to doubt fact A? As indicated in 3.2.2, such a question can be traced back to the more general acceptability criterion. However, as the question is formulated so specifically that it applies only to the argument from example, I consider it a question belonging to the category ‘specific to the argument from example’ (I).

(II) Generally applicable to all argument schemes: relevance criterion

For the argument from example, it has been proposed that the example should be relevant to the conclusion (see Table 4.5). Freeley and Steinberg (2000), for instance, write that “Advocates should determine whether the cases offered are relevant to the matter under consideration” (p. 154). Warnick and Inch (1989) are more specific and indicate the condition the examples should meet to be relevant to the conclusion: “To be relevant the examples must belong to the class of objects about which the claim is made” (p. 107). In other words, the examples should be selected from the group of people or things the conclusion applies to.

It is open to argument whether the questions of relevance, expressed in the current terms, are specific to the argument from example. In my opinion, they are not, as the relevance criterion concerns other types of argumentation as well. If they would be formulated in more specific terms, e.g. ‘Does the instance fall within the category of instances the generalization ranges over?’ I would consider them (more) unique for the argument from example.

Table 4.5 Evaluation questions falling under the category of (II) Relevance criterion.

<i>Evaluation question</i>	<i>Source</i>
(II) Relevance criterion	
Are the examples relevant to the claim of generality?	Warnick and Inch (1989, p. 107)
Does the example support the general claim it is supposed to be an instance of?	Walton (1996, p. 50)

<sup>85</sup> In Dutch: “Zijn er redenen om aan de juistheid van de voorbeelden te twijfelen?” (1994, p. 136).

Are the examples germane?	Reinard (1991, p. 196)
Is the example relevant?	Freeley and Steinberg (2000, p. 154)

#### 4.2.3 Conclusion and discussion

In argumentation theory, critical questions have been formulated that aim to evaluate the argument from example to a generalization. These questions can be classified according to six main categories: (I.1) the typicality/representativeness criterion, (I.2) the number of examples criterion, (I.3) the counterexamples criterion, (I.4) the strength of conclusion criterion, (I.5) the accuracy of the example criterion and (II) the relevance criterion. I argued that the first five categories of questions should be regarded as specific to the argument from example, whereas the last category should be considered as applicable to other types of argument as well. It should also be noted that in the present formulations of the evaluation questions, it is not always clear how the notions of typicality and representativeness should be conceived.

It can be argued that for the argument from example to a causal generalization, it is reasonable to distinguish between cause relevance and effect relevance, similar to the argument from cause to effect (see 3.2.3). For the argument from cause to effect, it sounds reasonable to require that the cause in the conclusion (Ai) is a clear member of the kind of causes (A) presented in the argument and also that the effect in the conclusion (Bi) is a clear member of the kind of effects (B) presented in the argument. For the argument from example to a causal generalization the principle is similar, but goes in a different direction: it sounds reasonable to require that the cause in the argument (Ai) is a clear member of the kind of causes (A) presented in the conclusion and also that the effect (Bi) in the argument is a clear member of the kind of effects (B) in the conclusion.

The requirements for quantitative and qualitative sampling, as described in 4.1.2, are obviously reflected in the argumentation-theoretical norms for evaluating the argument from example. More specifically, criteria referring to (I.1) typicality/representativeness and (I.2) the number of examples can be directly related to common requirements in sampling and scientific generalization. Other common principles in sampling theory, like the need for diversity and the criterion of random selection, are not included in the argumentation-

theoretical criteria. It appears reasonable, however, to demand that multiple examples are heterogeneous.

How about the categories (I.3) counterexamples criterion and (I.4) strength of conclusion criterion? In logic, a counterexample is an exception to a proposed false universal claim (e.g. "all visitors are bored"). So, counterexamples are considered to be important tools for evaluating the plausibility of universal assertions. In argumentation theory, that view has probably been adopted. However, unlike logic, argumentation theory has made insufficiently clear that the strength of the counterexample is dependent on the generality (or particularity) of the claim (exceptions are Schellens, 1985; Walton, 1996). Furthermore, overgeneralization, or drawing a conclusion that is too general, is considered a common reasoning fallacy (Walton, 1996, 1999). So, as expected, argumentation theory points to the risk of too certain or too strong a generalization.

### **4.3 Laymen criteria**

In section 4.2, argumentation-theoretical conceptions of the quality of the argument from example have been reported. In this section, the central question is: what criteria are used by laymen for the evaluation of the argument from example? Next, the method to address this question will be discussed.

#### **4.3.1 Method**

The method was the same as the one used in chapter 3, except for material and respondents.

#### **Material**

The material consisted of an open case and a closed case. Individual interviews, in which respondents are stimulated by an open and a closed case to reflect on norms for argument quality, appeared to be the most effective method to uncover laymen criteria (see chapter 2 and Šorm, Timmers & Schellens, 2007).

*Open case (appendix 4A-B).* In the open case, respondents were asked to imagine themselves in a job interview applying for a job as a speechwriter for a certain ministry (the name of the ministry was dependent on the claim). During the interview, they received an assignment to test whether they could write good speeches. The

assignment was to write down one strong and one weak argument supporting a claim in one of the minister's speeches. The claim to be supported was either (1) "If children are read to from childhood, they will be more proficient in the Dutch language" (in Dutch: "Als kinderen van jongs af aan worden voorgelezen, dan zullen ze de Nederlandse taal beter beheersen") or (2) "If employers contribute more to childcare, employees will be able to combine work and care more easily" (in Dutch: "Als werkgevers meer bijdragen aan de kinderopvang, dan zullen werknemers werk en zorg gemakkelijker kunnen combineren"). I chose these claims, because they were both causal generalizations that could potentially serve as premises in pragmatic argumentation. The respondents were told that the interviewer would ask them: why is one argument stronger than the other?

To increase the chance that the respondents would come up with the intended argument type, the respondents were told that they "had decided to support the minister's claim with examples". Furthermore, the respondents filled in a sentence leading to the right argument type: "Because I know a case in which that held true:..." (in Dutch: "Want ik ken wel een geval waarin dat opging:..."). After the respondents had written down their arguments, the interviewer asked them to read their arguments aloud and to explain why one argument was stronger than the other. The explanation allowed for identifying the respondents' criteria for evaluating the argument from example.

*Closed case (Appendix 4C-D).* The closed case consisted of a claim and a list of six supporting arguments. The claim was either (3) "If you strive for a healthy lifestyle, you will reduce the chance of illness" (in Dutch: "Als je een gezonde leefstijl nastreeft, dan verklein je de kans op ziektes") or (4) "If you choose an exact science as a subject, you will draw a high salary in future" (in Dutch: "Als je een exacte studierichting kiest, dan verdien je later een hoog salaris"). Again, I chose causal generalizations that could be used in pragmatic argumentation. Each claim was introduced by a few sentences about the claim's topic. For instance, as an introduction to the claim about exact science, it was stated that the government stimulates pupils to choose an exact science due to an expected shortage of science students and technicians in 2010.

The first of the six supporting arguments was supposed to be strong. The next five supporting arguments were considered weak as they violated one or two normative evaluation criteria found in argumentation theory (see also 4.2.2). The evaluation criteria that I

aimed to violate were the typicality/representativeness criterion, the number of examples criterion, the example accuracy criterion and the relevance criterion, all suggested in argumentation theory for the argument from example to a generalization. I specified the relevance criterion further into (a) cause relevance and (b) effect relevance. Although this distinction had not been made explicitly in argumentation theory, I argued that such a distinction would be reasonable (see 4.2.3). So, I wanted to know whether in practice people pay attention to relevance and if they do, whether they focus on the relevance of the action, the relevance of the effect or on both. I did not use the counterexamples criterion and the strength of conclusion criterion proposed in argumentation theory, as it proved impossible to adequately manipulate the strength of the arguments by means of these two criteria. Instead, I used the diversity criterion. This criterion has not been proposed in argumentation theory for the argument from example (except for Windes & Hastings, 1969), but it is nevertheless a common principle in sampling theory (see 4.1.2). In addition, making the premises less diverse appeared to be an easy way to weaken an argument. Table 4.6 shows the evaluation criteria that the arguments were supposed to violate.

Table 4.6 Evaluation criteria violated in closed-case arguments (translated from Dutch).

		<i>Claim (3)</i>	<i>Claim (4)</i>
		If you strive for a healthy lifestyle, you will reduce the chance of illness.	If you choose an exact science as a subject, you will draw a high salary in future.
<i>Criteria violated</i>		<i>Argument</i>	<i>Argument</i>
Strong, no criteria violated <sup>86</sup>	A	Simon de Wit started drinking less alcohol and once again has a well-functioning liver. Bart van Maanen quit smoking and now has less trouble with his bronchial tubes. Sander Schaafstal started to exercise more, and he now feels far more vital.	Geert-Jan van Harmelen studied dental surgery and now earns 5500 Euros a month. Bert Linthorst studied computer science and earns 3000 Euros a month. Jasper Simons studied architecture and now has an income of 4500 Euros a month.
Weak,	B	Theo van Gogh <sup>87</sup> (director,	Philip Freriks <sup>88</sup> , presenter of the

<sup>86</sup> After a re-examination of the material, I noticed that the argument (A) could be valued relatively low in diversity, because Simon, Bart and Sander are all male – unlike Maria, Edgar and Frederieke in argument (F).

<sup>87</sup> Theo van Gogh (1957-2004), a famous figure in the Netherlands, was a notorious chain smoker, heavy drinker and drug user. Therefore, the example saying that he lived healthily can be regarded as inaccurate.



violates: (1) example accuracy criterion (2) number of examples criterion		television producer and columnist) lived healthily and he had built up a fierce resistance to illnesses.	NOS news, studied chemistry and is in the top ten of the richest Dutchmen.
Weak, violates: (1) cause relevance criterion (2) number of examples criterion	C	Hanneke de Waal has had stomach reduction surgery and now she is less troubled by aching joints.	Janneke Oorthuys did a psychology degree and earns an above-average income.
Weak, violates: (1) effect relevance criterion (2) numbers of examples criterion	D	Robert de Graaf started drinking more water and his skin now looks less sallow.	Peter Philips studied veterinary medicine and has a nice job.
Weak, violates: (1) typicality/ repres. criterion (2) number of examples criterion	E	Robin van Persie <sup>89</sup> eats a balanced diet and is therefore in good condition.	The managing director of Shell studied physics and now earns millions of Euros a year.
Weak, violates: (1) diversity criterion	F	Maria de Graaf started to eat less fat, Edgar Ritmeester eats more fibre and grain than he used to, and Frederieke van Loon now eats more vegetables and fruit than before. All of them get sick less often than they used to.	Erna Bruin studied medical science, Paul Blok studied pharmacy and Renske Atsma studied biomedical sciences. They earn 4000 Euros a month.

<sup>88</sup> Philip Freriks, a well-known Dutch journalist and 8 o'clock News presenter, studied political sciences and never reached the top 10 of richest Dutchmen. So, the example can be considered inaccurate.

<sup>89</sup> Robin van Persie is a well-known Dutch football player who plays for the Dutch national team and has an international career.

The respondents were asked to put the arguments in order of strength, with the strongest argument in the first place and the weakest in the sixth place. After the respondents had ranked the arguments, the interviewer asked why a particular argument was put first, why another argument was put second and so on. Criteria were derived on the basis of their explanations.

### **Respondents**

Twenty respondents were interviewed. All respondents were visitors to a public library in Nijmegen, the Netherlands. Each of the respondents was paid ten Euros for participating. Among them, eleven were male and nine were female. The ages of the respondents varied from 17 to 62, and the average age was 31.7. The level of education varied from university education ('WO') to preparatory vocational education ('VMBO'). The majority (12) was on the level of pre-university education ('VWO') and higher vocational education ('HBO'). None of the respondents had received any formal training in the field of argumentation theory.

### **Booklet, Design, Procedure and Data-analysis**

Booklet, design, procedure, and data-analysis were similar to those of the study described in the previous chapter (see 3.3.1).

### **Interrater agreement**

To assess the extent to which fragments could be reliably related to criteria, interrater agreement was determined. The second rater was a colleague working at the Department of Business Communication Studies of the Radboud University Nijmegen (the Netherlands), who has special knowledge in the field of argumentation. First, she received a written task instruction. It stated that she would judge a number of fragments, in which phrases were marked. In addition, it explained that her task would be to label the phrases with criteria. The rater received a list of possible criteria she could choose from. Each criterion on the list was defined. Furthermore, the rater received the closed cases used in the interviews.

Five training fragments were presented, to give the second rater the opportunity to get familiar with the instrument and to ask for clarification. Afterwards, 11 randomly selected test fragments were presented. The results showed that the Kappa Measure of Agreement

value was .634 with a significance of  $p < .0005$  ( $N = 15$ ). According to Peat and Barton (2005, p. 268), this value generally represents good agreement.

#### 4.3.2 Results

Next, I will discuss the laymen criteria for evaluating the argument from example. First, the results of the open case will be reported, followed by the results of the closed case.

##### Interviews open case

It was first determined whether or not the arguments that the respondents had provided as a response to the open case were arguments from example to a causal generalization. The desired argument scheme underlying the arguments was 'In example A, measure M led to effect E'. Although the respondents were stimulated to use this desired scheme (see 'Material' in 4.3.1), none of the respondents produced both a strong and a weak argument from example according to this particular scheme. As a matter of fact, respondents were inclined to produce an argument in favor of the proposed action and an argument against the proposed action (instead of a strong and weak argument from example in favor of the claim), or to present other types of argument.

However, the interviews of six respondents can still be considered relevant to this study. Four respondents contrasted a generalization about multiple cases (as a strong argument) with a single case (as a weak argument). In explaining the qualitative difference between the two arguments, all four respondents referred to the importance of the *number of examples* upon which the conclusion is based. Three of them also referred to the *coincidence* that might be involved in a single case and one also mentioned the possibility of refuting by a *counterexample*. Here I present three fragments of interviews (held with different respondents) to illustrate these evaluation criteria. In each fragment, the strong argument is indicated by '(+)', the weak argument by '(-)' and the explanation about why one argument is stronger than the other by '(E)'. The claim that was to be supported by the arguments is also given. Consider the first fragment:

Fragment (1)
--------------

Claim to be supported:

If children are read to from childhood, they will be more proficient in the Dutch

---

language.

In English:

R: (...) as the strong argument I wrote down: research shows that children who are often read to will read more books in future so they will be better at spelling, better in sentence structuring and er stuff like that. (+ no. 1)

I: Hmmhmm.

R: An example, I can serve as an example myself, my mother has often read to me and now I like to read books myself. (+ no. 2)

I: Yes ...

R: So that applies, like, more when speaking of a large group. (E)

I: Yes, and what have you got for argument B, the weak argument?

R: The weak argument I thought up, my aunt never read to anyone and now my cousin is bad at Dutch. (-)

I: Okay, can you explain er, why argument A [the strong argument marked by '+ no. 1'] is stronger in supporting the standpoint than argument B [the weak argument marked by '-']?

R: Er yes, argument A [the strong argument marked by '+ no. 1'], suppose that it, that it is the case, right, that research has shown that. Then you have something er that is about a large group of people, and a conclusion is drawn from that, so that one is much better supported. And if you say in the case of argument B [the weak argument marked by '-'], my aunt never read to someone, so now my cousin is bad at spelling or something, that can also simply be a coincidence... (E)

In Dutch:

R: (...) bij het sterke argument had ik opgeschreven: uit onderzoek is gebleken dat kinderen die vaak worden voorgelezen ook zelf later meer boeken gaan lezen waardoor ze ook beter in spelling zullen zijn, beter in zinsbouw en dat soort ehh dingen. (+ no. 1)

I: Hmmhmm.

R: Een voorbeeld daarvan, kan ik zelf als voorbeeld staan, mijn moeder heeft veel voorgelezen en nou lees ik graag zelf boeken. (+ no. 2)

I: Ja ...

R: Dus dat is dan meer over een grote groep gesproken zeg maar. (E)

I: Ja, en wat heb je bij argument B, het zwakke argument?

R: Het zwakke argument heb ik bedacht, mijn tante las nooit voor en nu is mijn neefje slecht in Nederlands. (-)

I: Oké, kun je uitleggen ehm, waarom argument A [het sterke argument aangegeven met '+ no. 1'] sterker is ter ondersteuning van het standpunt dan argument B [het zwakke argument aangegeven met '-']?

R: Ehm ja, argument A [het sterke argument aangegeven met '+ no. 1'], stel dat het, dat het zo is hè, dat uit onderzoek is uitgebleken. Dan heb je iets ehm dat over een grote groep mensen gaat, waar wel een conclusie over getrokken is, dus die is veel beter onderbouwd. En als je gaat zeggen bij argument B [het zwakke argument aangegeven met '-'], mijn tante las nooit voor, dus nu is mijn neefje slecht in de spelling ofzo, dat kan ook gewoon toeval zijn ... (E)

---

In fragment (1), the respondent gives two strong arguments. The first strong argument (+ no. 1) is in fact an argument from authority: the respondent cites a certain study in which an inductive generalization was made. Then, another strong argument is given (+ no. 2) that can

be classified as an argument from example to a causal generalization. The weak argument (-) appears to be a variant of the desired argument scheme: in the particular case mentioned, the measure as proposed in the claim (reading from childhood) was not taken and the effect E (more proficiency) did not occur. In explaining (E) the qualitative difference between the arguments, the respondent refers to the inductive generalization in the first strong argument (+ no. 1) and suggests that the number of cases determines argument quality and that if only one case is cited, coincidence might play a role. In the next fragment, another respondent appears to apply the same criteria:

Fragment (2)

Claim to be supported:

If children are read to from childhood, they will be more proficient in the Dutch language.

In English:

R: Okay, as a er strong argument I wrote that children used to be read to more often and that they had a more thorough command of Dutch back then. (+)

I: Yes.

R: And as a weaker argument I have filled in that my parents used to read to me often and that I'm good at Dutch now. (-)

I: Okay, thank you. Er, can you explain why argument A is stronger in supporting the standpoint than argument B?

R: Well er, the most important thing is that er argument 1 is about a very large group of people and argument 2 about only one case and that er the chance that there is a coincidence, is bigger for the second argument. (E)

In Dutch:

R: Oké, als eh sterk argument heb ik ingevuld dat vroeger kinderen meer werden voorgelezen en dat ze toen het Nederlands ook beter beheersten. (+)

I: Ja.

R: En als zwakker argument heb ik ingevuld dat mijn ouders mij vaak voorlezen en dat ik nu goed Nederlands kan. (-)

I: Oké, dankjewel. Eh kun je uitleggen waarom argument A sterker is ter ondersteuning van het standpunt dan argument B?

R: Nou ehm, het belangrijkste punt is dat eh argument 1 over een heel grote groep mensen gaat en argument 2 over slechts één geval en dat eh de kans dat het van, toeval gewoon sprake is, groter is bij het tweede argument. (E)

In the following fragment (3), the respondent suggests that by using a single instance, a counterexample may prove that the claim is false:

Fragment (3)

Claim to be supported:

If children are read to from childhood, they will be more proficient in the Dutch language.

---

In English:

I: Can you read aloud first what you selected as the strong argument?

R: Er children who have been adopted, from a different country, and who are raised in a Dutch family, achieve a thorough command of the language. (+)

I: Yes, okay. What do you have as the weak argument?

R: Er, reading a story on a regular basis stimulated my child to tell the story word for word in the same way, including intonation and sentence structure. (-)

I: (...) Er why is one argument stronger than the other?

(...)

R: And er in my personal situation it is also a confirmation of, well that defi... or that theory, that it does work like that, because I can see in the case of my one child that it er by reading aloud regularly, er little children copy and imitate things, that that works as an example and that he copies it. But because it is a personal argument, I can also refute it, because another child er picks it up in a different way and is influenced a lot less with respect to language, in spite of er reading aloud a lot, reading. So there is also an individual difference between that. (E)

I: Hmmhmm.

R: And if you make a general theory, is a, you know broader proof, right, of many more cases in which that occurs you can then draw a conclusion more quickly, than if you simply, if there is just one person who, you know, experienced it. (E)

In Dutch:

I: Kunt u misschien eerst even voorlezen wat u bij het sterke argument heeft?

R: Ehm kinderen die geadopteerd zijn, uit een ander land dus, en in een Nederlands gezin opgroeien, leren de taal goed beheersen. (+)

I: Ja, oké. Wat heeft u bij het zwakke argument?

R: Ehm, regelmatig een verhaal voorlezen stimuleerde mijn kind om het verhaal op dezelfde wijze inclusief intonatie en zinsbouw woordelijk te vertellen. (-)

I: (...) Eh waarom het ene argument sterker is dan het andere argument?

(...)

R: En ehh in mijn persoonlijke geval is dat dan nog eens een bevestiging van, nou die defi... of die theorie, dat dat dus inderdaad ook zo werkt, omdat ik dan aan mijn eigen kind kan zien dat het ehh door regelmatig voor te lezen, ehh kleine kinderen kopiëren en doen dingen na, dat dat dus als voorbeeld werkt en dat hij het overneemt. Maar omdat het een persoonlijk argument is, ik kan het ook weerleggen, omdat een ander kind ehh dat op een andere manier oppikt en veel minder in de taal beïnvloed is, ondanks het ehh het vele voorlezen, lezen. Dus er zit op zich ook nog een individueel verschil tussen. (E)

I: Hmmhmm.

R: En als je een algemene theorie maakt, is een, zeg maar een breder ehh bewijs, hè, van veel meer gevallen waarin dat voorkomt kun je dan eerder een conclusie uit trekken, dan als je het gewoon, als er maar één persoon is die het meemaakt, zeg maar. (E)

---

In the strong argument (+), a group of children (adopted children) is mentioned that exhibits the causal relation in the claim, although it is not made explicit that this group is read to from childhood. This strong argument is contrasted with the weak argument (-) in which one case is mentioned. Like in fragment (1) en (2), the respondent clarifies (E)

that the number of cases determines argument quality, but here he or she also indicates that by using one example, the claim can be challenged by a counterexample.

Two respondents do not compare a generalization with a single instance. In the next fragment (4), two instances are presented in the strong argument and one instance in the weak argument:

Fragment (4)
--------------

Claim to be supported:

If children are read to from childhood, they will be more proficient in the Dutch language.

In English:

R: Er, for the strong argument I have er, because I know a case in which this held true. My sister had more time to read to her eldest. However, she had more tasks after her youngest was born. The eldest started speaking in full sentences earlier than the youngest. (+)

I: Okay.

R: And for the weak argument, because I know a case in which this held true, I was never read to and I fell behind in linguistic development and I needed a lot of coaching, Dutch. (-)

I: Okay. Can you explain why one argument is stronger than the other?

R: The first argument is stronger because it er, because it to, because you can, because you have two cases that you can compare with each other. So, it can be tested. The second er cannot be tested that easily and the question is er, it is a weaker argument because [it] does not always have to do with needing to be coached, because you haven't been read to, that correlation is not always clear.

(E)

In Dutch:

R: Ehh, bij het sterke argument heb ik ehh, want ik ken een geval waarin opging dat mijn zus heeft meer tijd gehad haar oudste voor te lezen. Ze kreeg echter meer taken nadat haar jongste werd geboren. De oudste maakt sneller volzinnen dan de jongste. (+)

I: Oké.

R: En bij het zwakke argument, want ik ken een geval waarin opging, ik ben zelf nooit voorgelezen en heb in mijn taalontwikkeling een achterstand opgelopen en moest veel bijlessen, Nederlands. (-)

I: Oké. Kunt u uitleggen waarom het ene argument sterker is dan het andere?

R: Het eerste argument is sterker omdat het ehh, omdat het te, omdat je het, omdat je twee gevallen hebt die je aan elkaar kan vergelijken. Dus, het is toetsbaar. Het tweede ehh is niet zo goed toetsbaar en de vraag is ehh, het is een zwakker argument omdat [het] niet altijd te maken heeft of dat je bijles moet hebben, omdat je niet voorgelezen bent, want die correlatie is niet duidelijk. (E)

As a strong argument (+), the respondent presents two cases. In one case, more reading led to better proficiency, whereas in the second case, less reading led to less proficiency. As a weak argument (-)

only one case is presented, in which a child (the respondent him- or herself) was not read to and got behind in linguistic development. The respondent then explains (E) that by comparing the two individual cases in the strong argument, it can be tested whether or not a *correlation* exists between two events, whereas on the basis of one case someone cannot be sure whether or not two events are related by correlation. Last, consider fragment (5):

Fragment (5)
--------------

Claim to be supported:

If employers contribute more to childcare, employees will be able to combine work and care more easily.

In English:

R: Okay, er, the strong argument, because I know a case in which that held true. The woman next door has three children. Besides doing her job she of course has to take care of her children. That is difficult in her situation, because she cannot do this at the same time. (+)

I: Okay.

R: The weak argument er, my neighbour owes money to various people. Therefore, he has to work very long at the office and he has no time to take care of his children. (-)

I: Thank you. Can you explain why A is stronger?

R: Er, well it, er, it is about a woman next door who has to work and she also has children and er, be, and, yeah it is stronger, because for the weak argument, er it is, that is more a very personal reason why that neighbour cannot take care of his children and work at the same time. Whereas, for the women next door, it is more of a general problem in society, so I think, because one is somewhat more personal than the other, yeah. (E)

I: Okay, and what do you mean exactly by er more a general problem?

R: Er, I think that a lot of people, a lot of families with children have this problem.

(E)

(...)

I: And what did you mean exactly by personal?

R: Er, well that it does not hold true for everybody and that it is really er the problem for the person, the individual himself.

I: Yeah, okay.

R: And therefore er not representative of others. (E)



In Dutch:

R: Oké, eh, het sterke argument, want ik ken wel een geval waarin dat opging. Mijn buurvrouw heeft drie kinderen. Naast haar werk moet ze natuurlijk ook voor haar kinderen zorgen. Dit gaat in haar situatie moeilijk, omdat ze dit niet tegelijk kan doen. (+)

I: Oké.

R: Het zwakke argument eh, mijn buurman heeft schulden bij verschillende kennissen. Hij moet daarom erg veel lange dagen maken op werk en heeft geen tijd om voor zijn kinderen te zorgen. (-)

I: Dankjewel. Kun je uitleggen waarom A sterker is?

R: Ehm, nou het, eh, het gaat over een buurvrouw die moet werken en ze heeft ook kinderen en eh, om, en, ja het is sterker, omdat bij het zwakke argument, eh het is, dat is meer een heel erg persoonlijke reden waarom die buurman niet voor zijn kinderen kan zorgen en werken tegelijk. Terwijl, bij de buurvrouw, dat is meer een algemeen probleem in de samenleving, dus ik denk, omdat de ene wat persoonlijker is dan de andere, ja. (E)

I: Oké, en wat bedoel je met eh meer een algemeen probleem precies?

R: Ehm, ik denk dat heel veel mensen, heel veel gezinnen met kinderen met dit probleem eh zitten. (E)

(...)

I: En wat bedoelde je precies met persoonlijk?

R: Eh, nou dat het niet voor iedereen geldt en dat het echt eh het probleem bij de persoon, de individu zelf zit. (E)

I: Ja ja.

R: En dus eh niet representatief is voor anderen. (E)

---

The respondent in fragment (5) gives examples of people having a hard time combining work and care, whereas the open case actually aimed at cases exhibiting the causal relation stated in the claim. So, the arguments appear to support a descriptive generalization (that is *not* stated in the claim to be supported) instead of a causal generalization (that *is* actually stated in the claim). Then the respondent explains that one argument is weaker than the other, because the example in the weak argument is not *representative* of others.

In sum, the following evaluation criteria were drawn from the interviews about the open case: criteria concerning the *number of examples*, *coincidence*, *counterexamples*, *correlation* between events and *representativeness*.

### Interviews closed case

The evaluation questions that have been found among the respondents can be grouped into (I) questions specific to the argument from example and (II) questions that are specific to a different argument scheme. Criteria that do not occur in theory are in *Italics*.

(I) Specific to the argument from example

Typicality criterion

Some respondents (15%) indicated that they found an argument relatively weak because an exceptional case was presented. This is illustrated by the next fragments:

Fragment (6)

Evaluated argument (E):

The managing director of Shell studied physics and now earns millions of Euros a year.

Supported claim:

If you choose an exact science as a subject, you will draw a high salary in future.

In English:

I: Okay, and argument E is in position 3, so I see. Can you explain that?

R: Yes because er, it is, you know, an extreme case, so not er, not so you know, important, or yeah it is, there is a big chance that it is simply, like, an exception, and not er, not a logical consequence of the fact that you're going to do an exact science.

In Dutch:

I: Oké, en argument E staat op plaats 3 zie ik. Kun je dat uitleggen?

R: Ja want ehm, het is wel zeg maar een extreem geval, dus niet zo ehh, niet zo belangrijk zeg maar, of ja het is, het is een grote kans dat het gewoon een uitzondering is zeg maar, en niet ehh, niet een logisch gevolg van dat je een exacte studie gaat doen.

Fragment (7)

Evaluated argument (E):

The managing director of Shell studied physics and now earns millions of Euros a year.

Supported claim:

If you choose an exact science as a subject, you will draw a high salary in future.

In English:

I: You found this stronger than argument E? [inaudible] position 2.

R: Yes, er in general er, the director of Shell studied physics and now earns millions of Euros. Er so if you choose an exact science, then you will draw a high salary in the future. Er how many directors of Shell are there actually and how is his study er in physics directly connected to his high salary? In general, if you studied physics, you indeed have a, a good, well, a considerable analytical ability I think, that I definitely believe. You also see indeed that executives have often done exact sciences already. He could just as easily have done economics, or maybe law and maybe he's just a very clever guy, who at that, at that moment didn't know exactly in what direction he wanted to go. So er I find it an exception.

In Dutch:

I: U vond hem sterker dan argument E? [onverstaanbaar] plek 2.

R: Ja, ehh over het algemeen ehh, de directeur van Shell heeft natuurkunde gestudeerd en verdient nu miljoenen euro's. Ehh dus als je een exacte studierichting kiest, dan verdien je later een hoog salaris. Ehh hoeveel directeuren van Shell zijn er eigenlijk en wat heeft zijn studie ehh natuurkunde nu rechtstreeks verband te houden met zijn hoge salaris? Over het algemeen, als je natuurkunde hebt gestudeerd, dan heb je inderdaad wel een, een goeie, nou, een behoorlijk analytisch vermogen denk ik, dat geloof ik zeker. Je ziet ook inderdaad van topmensen hebben vaak al inderdaad exacte studierichtingen gedaan. Voor hetzelfde geld had hij economie gedaan, of misschien rechten en is het gewoon een slimme kerel inderdaad, die op dat, in dat moment nog niet precies weet welke richting die op wil gaan. Dus ehh ik vind het een uitzondering.

---

#### Number of examples criterion

The results show that laymen regard the number of examples as an indication of argument quality (45%). Some simply explain that giving more examples is better, but others explain what it is exactly that makes the number of examples important. The following respondent, for instance, explains that a single example might be an exception that accidentally justifies the conclusion:

Fragment (8)
--------------

Evaluated argument (F):

Maria de Graaf started to eat less fat, Edgar Ritmeester eats more fibre and grain than he used to, and Frederieke van Loon now eats more vegetables and fruit than before. All of them get sick less often than they used to.

Supported claim:

If you strive for a healthy lifestyle, you will reduce the chance of illness.

In English:

I: Next, I see that you put argument F in second place...

R: Yes.

I: Can you explain why?

R: (...) And it is also about more people, so I found that rather good. That one and that one and that one, they stopped eating and now they have fewer diseases.

I: And why did you find that good, more people?

R: Er, yes, the same in the previous case, that you think like, if you say one person has started to eat better and now he has fewer diseases, then you can think, yeah, gosh nice, that can also be a simple coincidence. But if you say that one and that one and that one, then I think after all, gosh it may be right after all. So that's why I put that one in second place.

In Dutch:

I: Dan zie ik dat je op de tweede plaats argument F hebt gezet...

R: Ja.

I: Kun je uitleggen waarom?

R: (...) En het gaat ook over meerdere mensen, dus dat vond ik wel goed. Die en die en die, zijn gestopt met eten en nou hebben ze minder kwaaltjes.

I: En waarom vond je dat goed, meerdere mensen?

R: Ehm, ja hetzelfde bij de vorige casus een beetje, dat je denkt van, als je zegt één persoon is beter gaan eten en die heeft nou minder kwaaltjes dan kan je denken, ja, goh leuk, dat kan ook gewoon toeval zijn. Maar als je zegt van die en die en die allemaal, dan denk ik toch, goh misschien zit er dan toch wel wat in. Dus die had ik daarom op de tweede plek gezet.

---

Another respondent relates the number of examples to the diversity of characteristics in the examples:

Fragment (9)
--------------

Evaluated argument (D):

Robert de Graaf started drinking more water and his skin now looks less sallow.

Supported claim:

If you strive for a healthy lifestyle, you will reduce the chance of illness.

In English:

I: Okay, yes. And then you put D in fourth place...

R: Yeah, yeah, yeah. Basically also a rather er, a er good one, but I don't find it all that strong. (...) Maybe also because it er, because one, one person is mentioned. With one change in his lifestyle and mentioning three persons, maybe that sounds a bit stronger.

I: And why does that sound stronger?

R: Don't know, probably because of three different aspects, that that, that that strengthens the claim a bit. More argumentation for why a healthy lifestyle reduces the chance of illness.

In Dutch:

I: Oké, ja. En dan heb je nog D op de vierde plaats ...

R: Tja, tja, ja. An sich ook wel een ehh, een ehh goeie, maar dat vind ik niet zo sterk. (...) Misschien ook omdat het ehh, omdat er één, één persoon wordt genoemd. Met één verandering in zijn leefpatroon en het noemen van drie personen, dat dat misschien wat sterker overkomt.

I: En waardoor komt dat sterker over?

R: Weet ik niet, waarschijnlijk door drie verschillende aspecten, dat dat, dat dat de stelling wat ehh sterker maakt. Meer beargumentatie waarom een gezonde leefstijl kans op ziektes verkleint.

---

The next respondent thinks that the more confirmations of the generalization are given, the higher the chance is that the generalization will be confirmed again:

Fragment (10)

Evaluated argument (F):

Maria de Graaf started to eat less fat, Edgar Ritmeester eats more fibre and grain than he used to, and Frederieke van Loon now eats more vegetables and fruit than before. All of them get sick less often than they used to.

Supported claim:

If you strive for a healthy lifestyle, you will reduce the chance of illness.

In English:

I: Okay, okay, because you put F in er second place, I see.

(...)

R: Yes, well actually also because there are more than the others, because for the rest, it is all one. For B, C, D and E.

I: Okay, and why do you find that to be a stronger argument, that there are more?

R: Er... yeah, because then there are more examples at once of, er, people, to whom that argument, you know, applies.

I: Hmmhmm, and why is that strong?

R: Er well yeah, the more people something holds true for, the more plausible it is that that holds true for you too.

In Dutch:

(I: Oké, oké, want je had F op ehh nummer 2 hè, zie ik.

(...)

R: Ja, nou eigenlijk ook omdat het er ehh meer zijn dan de anderen, want voor de rest zijn het allemaal één. Bij B, C, D en E.

I: Oké, en waarom vind je dat sterker, dat het er meer zijn?

R: Ehm ... ja, omdat het dan in één keer meer voorbeelden zijn van, ehh, mensen voor, voor wie dat argument geldt, zeg maar.

I: Hmmhmm, en waarom is dat sterk?

R: Ehh nou ja, voor hoe, hoe meer mensen iets geldt, ehh ja, hoe aannemelijker het is dat dat dan ook voor jou geldt.

And one person simply prefers more examples because it is more impressive:

Fragment (11)

Evaluated argument (A):

Simon de Wit started drinking less alcohol and once again has a well-functioning liver. Bart van Maanen quit smoking and now has less trouble with his bronchial tubes. Sander Schaafstal started to exercise more, and he now feels far more vital.

Supported claim:

If you strive for a healthy lifestyle, you will reduce the chance of illness.

In English:

I: But, you did find this one weaker than argument A.

R: Yes, yes. But maybe also because there is a lot of information in A and that it is therefore somewhat, somewhat more impressive, I don't know... yeah. In A you

see three things, you know, so those together perhaps lead to a bigger er bigger argument.

In Dutch:

I: Maar, je vond hem wel zwakker dan argument A, dat wel.

R: Ja, ja. Misschien ook omdat er bij A heel veel staat en dat daarom wat, wat imposanter overkomt, ik weet het niet ... ja. Bij A zie je drie dingen staan, zeg maar, die dus samen ook misschien zorgen voor een groter ehh groter argument.

---

It may be argued that this particular evaluation does not concern the content of the argument. After all, the respondent appears to use the number of examples criterion as a peripheral cue: 'the more information given, the stronger the argument'.

#### Example accuracy criterion

Laymen use criteria referring to the accuracy of the example (45%). As illustrated by the following fragment, some doubt if the example corresponds to reality:

Fragment (12)
---------------

Evaluated argument (B):

Theo van Gogh (director, television producer and columnist) lived healthily and he had built up a fierce resistance to illnesses.

Supported claim:

If you strive for a healthy lifestyle, you will reduce the chance of illness.

In English:

R: B is er last, because er should we know, we all remember him on TV, smoking and sweating and fat. So really he does not have such a healthy lifestyle. That taken into consideration, you have to have prior knowledge, if you don't know that man then er, it, then that argument would not belong under B, but I know this man and he was not healthy, did not have a healthy lifestyle. So yeah, he has not built up a good resistance, so that's why that one is on B.

I: That's why that one is in last place.

R: In, yeah in last place, yeah.

In Dutch:

R: B is ehh op het laatst, omdat ehh zouden wij weten, wij weten allemaal dat hij rokend en zwetend en dik op de tv zat. Dus echt zo gezond leeft die niet. Dat meegenomen, dus je moet wel voorkennis hebben, als je die meneer niet kent dan ehh schiet, dan hoort die dus niet op B, maar ik ken deze meneer en die had geen gezonde leeftijd, gezonde leefstijl. Dus ja, die heeft dus ook geen goede weerstand opgebouwd, dus daarom komt die op B.

I: Daarom komt die op de laatste.

R: Op, ja op de laatste plek, ja.

---

So, the respondent in fragment (12) questions if Theo van Gogh had a healthy lifestyle and built up fierce resistance to illnesses, as proposed in the example.

Another respondent doubts that Philip Freriks, the NOS-news presenter, studied chemistry:

Fragment (13)

Evaluated argument (B):

Philip Freriks, presenter of the NOS news, studied chemistry and is in the top ten of the richest Dutchmen.

Supported claim:

If you choose an exact science as a subject, you will draw a high salary in future.

In English:

I: Okay, and B?

R: B ...

I: What can you say about that?

R: Philip Freriks, well I, I, I think he is a linguistics guy. And I don't believe he, he studied chemistry.

In Dutch:

I: Oké, en B?

R: B ...

I: Wat kunt u daarover zeggen?

R: Philip Freriks, nou ik, ik, volgens mij is dat een talenkerel. En ik geloof niet dat hij scheikunde gestudeerd heb, heeft.

Concreteness criterion

The criterion of concreteness was also used by the respondents in this study (35%). Some explicitly mention the term 'concrete' in relation to argument quality, like the respondent in the next fragment does:

Fragment (14)

Evaluated argument (A):

Simon de Wit started drinking less alcohol and once again has a well-functioning liver. Bart van Maanen quit smoking and now has less trouble with his bronchial tubes. Sander Schaafstal started to exercise more, and he now feels far more vital.

Evaluated argument (E):

Robin van Persie eats a balanced diet and is therefore in good condition.

Supported claim:

If you strive for a healthy lifestyle, you will reduce the chance of illness.

In English:

R: Er, but why I find it [argument E] weaker than A, is because A er also gives concrete examples. Yeah, no, E does that too by the way.

In Dutch:

R: Ehm, maar waarom ik hem [argument E] zwakker vind dan A, is omdat A ehh geeft concrete voorbeelden ook. Ja, nee dat doet E trouwens ook.

---

Others just refer to concreteness, as the next fragment shows:

Fragment (15)
---------------

Evaluated argument (A):

Simon de Wit started drinking less alcohol and once again has a well-functioning liver. Bart van Maanen quit smoking and now has less trouble with his bronchial tubes. Sander Schaafstal started to exercise more, and he now feels far more vital.

Evaluated argument (F):

Maria de Graaf started to eat less fat, Edgar Ritmeester eats more fibre and grain than he used to, and Frederieke van Loon now eats more vegetables and fruit than before. All of them get sick less often than they used to.

Supported claim:

If you strive for a healthy lifestyle, you will reduce the chance of illness.

In English:

R: Er I have argument A on 1 (...) And each example also indicates what is good about it, you know, because Simon de Wit, less alcohol, a well-functioning liver. Bart van Manen, quit smoking, less trouble with his bronchial tubes, you know separately indicated. And that's why that one is before F, because in F it's also only about three people, but then it says only sick less often than they used to.

I: And why do you find that weaker? That it only says sick less often than they used to?

R: Er, I don't find the word sick all that convincing. Because, yeah, what is sickness then? And er, it says a lot less as a result.

I: Hmmhmmm, less than...?

R: Less than er in answer A where er it was indicated each time what was exactly good about it.

In Dutch:

R: Ehm ik heb argument A op 1 (...) En bij elk voorbeeld is er ook aangegeven wat er dan goed aan is, zeg maar, want Simon de Wit, minder alcohol, een goed functionerende lever. Bart van Manen, gestopt met roken, minder last van zijn luchtwegen, zeg maar los van elkaar aangegeven. En daarom staat die boven F, want bij F staat er alleen voor drie mensen ook, maar dan staat er alleen minder last van kwaaltjes.

I: En waarom vind je dat zwakker? Dat er alleen minder last van kwaaltjes staat.

R: Ehm, ik vind kwaaltjes als woord vind ik niet zo overtuigend. Want, ja, wat zijn kwaaltjes dan? En ehm, daardoor zegt het ook veel minder.

I: Hmmhmm, minder dan ...?

R: Minder dan ehh bij antwoord A waarbij ehh steeds werd aangegeven wat er precies goed is.

---



It is possible that the concreteness-question is asked prior to other evaluation questions, like the question of relevance. After all, if an instance is not clear-cut, it will be more difficult or even impossible to judge whether or not and to what degree it belongs to a category. For instance, the respondent in fragment (15) could have been unable to judge if 'sickness' is a good member of the category 'illness', as he finds it unclear what this 'sickness' precisely is.

### Causality criterion

Laymen question the causality suggested in the example (85%). When they do so, they sometimes refer to necessity and sometimes to sufficiency. If someone refers to necessity, then it is asked whether the effect would have occurred in the absence of the action. If someone refers to sufficiency, it is questioned whether the occurrence of the action brings about the consequence<sup>90</sup>. In the following fragment, the respondent refers to necessity:

Fragment (16)
---------------

Evaluated argument (B):

Philip Freriks, presenter of the NOS news, studied chemistry and is in the top ten of the richest Dutchmen.

Supported claim:

If you choose an exact science as a subject, you will draw a high salary in future.

In English:

I: (...) and on to B, you put that one in second to last place.

R: Hmmhmm.

I: Can you explain why?

R: Yes, because, it seems to me that there is no real relation between studying chemistry and then ending up at the NOS-news, so er, I mean then I would have rather a study journala, journalism or er something in that direction, yeah, it doesn't seem necessary to me to study chemistry, so, yeah.

I: Okay.

R: I don't think that it's true that, just because he studied chemistry, he [is?] in the top 10 of richest Dutchmen then. He probably became a presenter coincidentally, so I don't find that one such a strong one, actually.

I: Okay.

R: Not really a logical connection.

---

<sup>90</sup> For a more extensive discussion about sufficiency and necessity, see chapter 3 about the argument from cause to effect.

In Dutch:

I: (...) en dan gaan we naar B, die heb je op de één na laatste plek.

R: Hmmhmm.

I: Kun je uitleggen waarom?

R: Ja, omdat, het lijkt me niet echt verband hebben dat je scheikunde studeert en dan bij het NOS-journaal terecht komt, dus ehh, ik bedoel dan zou ik eerder een studie journala, journalastiek of ehh iets in die richting, ja, lijkt me niet dat je daarvoor scheikunde gestudeerd moet hebben, dus, ja.

I: Oké.

R: Ik denk niet dat het zo is dat, omdat hij scheikunde heeft gestudeerd, dat hij dan in de top 10 van rijkste Nederlanders [staat?]. Hij is waarschijnlijk toevallig presentator geworden, dus dat vind ik eigenlijk niet zo'n sterke.

I: Oké.

R: Niet echt logisch verband.

---

So, here the respondent questions whether the action (studying chemistry) was necessary to reach the effect (being in the top 10 of richest Dutchmen). In the next fragment, the respondent refers to sufficiency:

Fragment (17)
---------------

Evaluated argument (E):

The managing director of Shell has studied physics and now earns millions of Euros a year.

Supported claim:

If you choose an exact science as a subject, you will draw a high salary in future.

In English:

I: Okay, and if you were to say something about the arguments separately, about E for example?

R: Arguments separately, the managing director of Shell studied physics and now earns millions of Euros a, a year. Well, that is nonsense of course, because physics does not guarantee millions of Euros a year. That is simply a false argument, or er simply, there are also people who studied physics simply like, who earn ten Euros a month, for instance.

In Dutch:

I: Oké, en als u iets moet zeggen over de argumenten afzonderlijk, over E bijvoorbeeld?

R: Argumenten afzonderlijk, de directeur van Shell heeft natuurkunde gestudeerd en verdient nu miljoenen per, euro's per jaar. Nou, dat is natuurlijk een onzin, want natuurkunde staat niet garant voor miljoenen euro's per jaar. Dat is gewoon een vals argument, van ehh gewoon, er zijn ook mensen die natuurkunde gestudeerd hebben gewoon van, die een tientje per maand verdienen, bij wijze van spreken.

---

So, the respondent in fragment (17) questions if the action (studying physics) is sufficient to bring about the effect (earning millions of Euros a year).

### Relevance criterion

Laymen pay attention to the relevance of the example to the claim (85%), so they decide whether or not the argument is connected to the matter under consideration. Some focus on the relevance of the action in the argument to the action in the claim, some on the relevance of the effect in the argument and some focus on the relevance of the argument as a whole. The next fragment illustrates a case in which someone focuses on the relevance of the effect in the argument:

Fragment (18)
---------------

Evaluated argument (D)

Peter Philips studied veterinary medicine and has a nice job.

Supported claim:

If you choose an exact science as a subject, you will draw a high salary in future.

In English:

I: Okay, I see that D is left, in the second to last place, argument D.

R: That is Peter Philips. Yes, he has a really nice job and er, so if you [choose?] an exact science, you will draw a high salary in the future, of course the... one thing does not have to do a damned thing with the other actually. He has a nice job, okay, but it tells me nothing about how much he earns.

I: And for that reason you find er weaker than...

R: Yeah... well if you choose an exact science, then you will draw a high salary in future, then you should not be using this one indeed. If you choose an exact science and then you get a really nice job, then that one would be perfect of course.

In Dutch:

I: Oké, D zie ik dan nog, op de één na laatste plek, argument D.

R: Dat is Peter Philips. Ja, die heeft een hele leuke baan en ehh, dus als je een exacte studierichting [kiest?], verdien je later een hoog salaris, natuurlijk een... heeft er eigenlijk, geen zak met elkaar te maken. Hij heeft een leuke baan, oké, maar dat zegt me niets over hoeveel hij verdient.

I: En om die reden vindt u ehm zwakker dan...

R: Ja... nou als je een exacte studierichting kiest, dan verdien je later een hoog salaris, dan zou je deze niet in moeten gebruiken inderdaad. Als je een exacte studierichting kiest en dan krijg je een ontzettend leuke baan, dan zou die natuurlijk perfect zijn.

It also appears that respondents in this study judge category membership before they determine the relevance of the argument. In other words, they appear to use the heuristic: 'if an instance in the argument falls within a claim category, then the argument is relevant'. The next fragment illustrates this phenomenon:

Fragment (19)

Evaluated argument (C):

Janneke Oorthuys did a psychology degree and earns an above-average income.

Supported claim:

If you choose an exact science as a subject, you will draw a high salary in future.

In English:

R: No, in the, yes, in last place I put er the one who did the psychology degree because yeah, that is simply not an exact science.

I: Hmmhmm.

R: And, yeah, that just doesn't have anything to do with it, you know.

I: It does not have anything to do with what?

R: With er exact sciences leading to a er high salary.

I: Hmmhmm.

R: Because it is simply not an exact science, so, yeah, then er, it is simply not an argument at all.

In Dutch:

R: Nee, op de, ja, op de laatste plaats had ik dus ehh degene die de psychologie-opleiding heeft gedaan want ja, dat is gewoon geen exacte studie.

I: Hmmhmm.

R: En, ja, dat heeft er gewoon niks mee te maken, zeg maar.

I: Waar heeft het niks mee te maken?

R: Met ehh dat exacte studies tot een ehm hoog salaris leiden.

I: Hmmhmm.

R: Want het gaat gewoon niet over een exacte studie, dus, ja, dan ehh, het is gewoon helemaal geen argument.

So, the respondent in fragment (19) suggests that 'Psychology' in the argument is not an instance of the category 'exact science' in the claim, so that the argument cannot be related to the claim.

Some refer to the *degree* of category membership and also to the *degree* of relevance. Consider the next fragments:

Fragment (20)

Evaluated argument (F):

Maria de Graaf started to eat less fat, Edgar Ritmeester eats more fibre and grain than he used to, and Frederieke van Loon now eats more vegetables and fruit than before. All of them get sick less often than they used to.

Supported claim:

If you strive for a healthy lifestyle, you will reduce the chance of illness.

In English:

I: Then I would like to know why you put F in first place.

R: Er, I put F in first place, because I think that that one is most er about a healthy lifestyle. Namely eating less fat, more fibre and grain and more vegetables and fruit. Those are all positive things, you know, for working on your health in a positive way. Whereas in the others it says a lot of er stopping with negative things

and this is er starting with positive things, you know.

In Dutch:

I: Dan wilde ik graag weten waarom je F op 1 hebt staan.

R: Ehh, ik heb F op 1 gezet, omdat ik vind dat daarin het meest ehh gaat over een gezonde leefstijl. Namelijk minder vet eten, meer vezels en granen en meer groenten en fruit. Dat zijn allemaal positieve dingen, zeg maar, voor op een positieve manier aan je gezondheid werken. Terwijl in de anderen staan veel ehh stoppen met negatieve dingen en dit is ehh beginnen met positieve dingen, zeg maar.

---

Fragment (21)
---------------

Evaluated argument (E):

The managing director of Shell studied physics and now earns millions of Euros a year.

Supported claim:

If you choose an exact science as a subject, you will draw a high salary in future.

In English:

I: Well, in er first place you put E, I see.

R: Yes.

I: Can you tell why you put that one in first place?

R: Yes, it er has to do with er, what does exact mean? What is an exact science? Well, in any case it is physics, chemistry in my opinion. Er computer science, for example architecture, and er in E it is the director of Shell, has studied physics ...

I: Hmmhmm.

R: ... and now earns millions of Euros. Er Shell is also an er, an er exact field, you know. I don't know what, what, what he does exactly as a director. But I assume that he got in Shell with physics and did something there and moved up eventually.

I: Hmmhmm.

R: So from then, then it connects to er an exact science. And I ranked them in that way, you know, the less er, exact it is, the less important it is as an argument.

In Dutch:

I: Nou, op ehh de eerste plek staat E, zie ik.

R: Ja.

I: Kun je vertellen waarom je die op de eerste plek hebt gezet?

R: Ja, dat ehh heeft met ehh, wat is exact? Wat is een exacte studierichting? Nou, het is in ieder geval natuurkunde, in mijn ogen scheikunde. Ehh informatica, bijvoorbeeld bouwkunde, en ehh bij E is het de directeur van Shell, heeft natuurkunde gestudeerd ...

I: Hmmhmm.

R: ... en verdient nu miljoenen euro's. Ehh Shell is ook nog een ehh, een ehh exact veld, zeg maar. Ik weet niet wat, wat, wat hij als directeur precies doet. Maar ik neem aan dat hij met natuurkunde binnen is gekomen bij Shell en daar wat heeft gedaan en uiteindelijk op is geklommen.

I: Hmmhmm.

R: Dus vanuit dan, dan sluit het aan bij ehh een exacte studierichting. En zo is het naar beneden gegaan zeg maar, hoe minder ehh, exact, hoe minder belangrijk als argument, zeg maar.

---

In fragment (20), the respondent states that the strongest argument is 'most' about a healthy lifestyle, suggesting that the single measures taken in that argument (eating less fat etc.) are relatively clear members of the category 'healthy lifestyle'. In fragment (21), the respondent suggests that the evaluation of the arguments is dependent on the degree to which the study named in the premise is a member of the category 'exact science' in the claim.

(II) Specific to a different argument scheme: *action desirability criterion*

In a few interviews (15%), respondents appeared to be concerned with the desirability of the action proposed in the argument. Some judgments are based on attitudes towards the action mentioned in the example, as the next fragment illustrates:

Fragment (22)
Evaluated argument (D): Robert de Graaf started drinking more water and his skin now looks less sallow.
Supported claim: If you strive for a healthy lifestyle, you will reduce the chance of illness.
In English: I: Hmmhmm, okay. And in fifth place you put argument D... R: Yes, I find that a lesser argument, I mean, drinking water is healthy, but... I find it less strong, less important. I think that it is more important to, to inform people on things like fish, like more vegetables and fruit, that way you know. Drinking water, that I think is a somewhat lesser, a less important argument.
In Dutch: I: Hmmhmm, oké. En op de vijfde plek staat argument D... R: Ja, dat vind ik ook wat minder, zo van water drinken is wel gezond, maar ... ik vind hem minder sterk, minder belangrijk. Ik denk dat het belangrijker is om, om mensen te informeren over dingen als vis, als meer groenten en fruit, op die manier zeg maar. En dan water drinken vind ik dan wel iets, een minder belangrijk argument.

So, in fragment (22) the argument is evaluated negatively, because the action of drinking water is considered relatively unimportant.

Table 4.7 shows the laymen criteria for the argument from example. Criteria that do not correspond to argumentation theory are in Italics.

Table 4.7 Laymen criteria for the argument from example (per type of stimulus material) and agreement among respondents (the number of closed case individual interviews in which the criterion was mentioned divided by the total number of closed case individual interviews) (laymen criteria that do not correspond to theory are in italics).

Nature of criterion	Criterion	Type of stimulus material		Agreement among respondents, based on closed case interviews
		Open case	Closed case	
Specific to the argument from example	Typicality	+	+	3 of 20 (15%)
	Number of examples	+	+	9 of 20 (45%)
	Counterexample	+	-	0 of 20 (0%)
	Example accuracy	-	+	9 of 20 (45%)
	<i>Concreteness</i>	-	+	7 of 20 (35%)
	<i>Causality</i>	-	+	17 of 20 (85%)
	Relevance (of the cause, of the effect or of the argument as a whole)	-	+	17 of 20 (85%)
Specific to a different argument scheme	<i>Action desirability</i>	-	+	3 of 20 (15%)

#### 4.4 Conclusion and discussion

In section 4.2, I examined the normative considerations on the quality of the argument from example in argumentation theory. In section 4.3, I uncovered laymen's considerations on the quality of the argument from example. The central question in this chapter was: to what extent do lay criteria for the argument from example correspond to the criteria formulated in argumentation theory?

A considerable number of laymen in this study judge the quality of the argument from example on the basis of the number of examples and on the basis of the accuracy of the example. This is in agreement with argumentation theory. A large majority of the laymen in this study use the relevance criterion. This criterion, however, is sometimes used on a more concrete level than is suggested in argumentation theory, as some focus on the relevance of the *cause* in

the argument and others on the relevance of the *effect* in the argument. In addition, it appears that category membership plays a role in relevance judgments: laymen in this study appear to determine whether, and to what degree, an instance falls into a conclusion category before they judge relevance. Only a small minority of laymen in this study focus on typicality, which means that they wonder if the case presented is an exceptional one involving atypical circumstances. Only one respondent mentioned the possibility of undermining generalizations by counterexamples, in the interview on the open case. This was to be expected considering the setup of the material: all generalizations were presumptive in nature (cf. Walton, 1999). If the respondents would have compared an argument from example resulting in a universal generalization (“e.g. *All* individuals striving for a healthy lifestyle will *at all times* reduce the chance of illness”), with the same argument resulting in a presumptive generalization (“Individuals striving for a healthy lifestyle will reduce the chance of illness”), for instance, they would have possibly been more triggered to refer to counterexamples.

In sum:

Corresponding criteria with relatively high agreement among respondents:

- The relevance criterion: the example should connect adequately to the conclusion.
- The number of examples criterion: there should be enough examples cited.
- The example accuracy criterion: the example should correspond to reality.

Corresponding criteria with relatively low agreement among respondents:

- The typicality criterion: the example should be typical of the kinds of cases the generalization ranges over.
- The counterexample criterion: there should not be a counterexample undermining the conclusion.

There are also differences between laymen criteria for the argument from example and theoretical criteria. Laymen did not refer to the strength of conclusion in their evaluations, as suggested in theory. This was, however, expected on the basis of the setup of the material, as the claims that were presented in the cases did not vary in this respect. A reasonable number of respondents used a concreteness norm; a norm that is familiar in argumentation theory<sup>91</sup>,

---

<sup>91</sup> The pragma-dialecticians, for example, refer to concreteness in their tenth commandment: “Discussants may not use any formulations that are



but that has not been included in evaluation criteria for the argument from example specifically. A large majority of laymen evaluated the argument from example on the basis of causality, something that is not suggested in argumentation theory for the argument from example. It appears justified that laymen do focus on causality, considering the fact that they judged causal generalizations supported by examples that were supposed to exhibit the causal relation in the claim. The theoretical criteria were aimed at the argument from example in general and not specifically at argumentation from example to a *causal generalization*. Therefore, the disagreement between laymen and argumentation theory could be expected. Last, a few laymen judged the desirability of the action in the argument from example, something they should not do according to argumentation theory. The pragmatic context may have triggered such a criterion. However, using this criterion appears to be irrelevant to this type of argument, because even if the action in the argument is not desirable, the causal generalization it supports may still hold. In sum, the non-corresponding criteria are:

Non-corresponding criterion that was not mentioned by respondents:

- The strength of conclusion criterion: the conclusion should not be too strongly formulated (theory only, as expected, because of material).

Non-corresponding criterion with relatively high agreement among respondents:

- The causality criterion: the examples should exhibit a causal relation (laymen only, as expected, because theory did not aim particularly at the argument from example to a causal generalization).

Non-corresponding criteria with relatively low agreement among respondents:

- The concreteness criterion: the examples should be concrete (laymen only, unexpected on the basis of theoretical scheme-dependent criteria).
- The action desirability criterion: the action in the example should be desirable (laymen only, unexpected, because irrelevant to the argument from example).

---

insufficiently clear or confusingly ambiguous" (Van Eemeren & Houtlosser, 2006, p. 18).

## Discussion

Laymen criteria for the evaluation of the argument from example correspond to a considerable extent to the theoretical criteria. About the corresponding criteria, however, it can be said that the relevance criterion, the number of examples criterion and the example accuracy criterion are agreed upon by a considerable number of laymen in this study, whereas the typicality criterion and the counterexample criterion are agreed upon by a relatively small number of laymen in this study.

The fact that most laymen in this study did not focus on counterexamples could be expected considering the set-up of the material, but the relatively low agreement on the typicality criterion was rather unexpected. It is possible that the causality in the material stimulated respondents to focus primarily on causality conditions and also on the relevance of the action and the effect in the argument. Had the material contained arguments from example to a descriptive generalization, respondents might have given more attention to typicality in the examples.

A reasonable number of laymen apply a concreteness-criterion, which was not advised in argumentation theory for the argument from example specifically. How can that result be explained? I already argued in 4.3.2 that concreteness may enable laymen to adequately judge whether instances in the premises are clear members of conclusion categories (and as a consequence, to adequately judge the relevance of the argument). Although from a normative point of view it is desirable for each argumentation that it is specific, it is quite natural that in practice, this requirement becomes extra important in the case of the argument from example, having instances and categories as its essential characteristics. Therefore, it should not come as a surprise that respondents in this study clearly focus on concreteness.

An alternative explanation is that respondents have not interpreted the arguments from example in the material as *argumentative*, but as *illustrative*. In that case, it is understandable that they wanted the examples to be concrete. After all, considering the fact that an illustrative example's (main) purpose is to put out something clearly or pictorially, it should be at least concrete.

This study gives reason to think that evaluation questions for the argument from example need some refinement. A few laymen in this study are concerned with typicality/representativeness as advised in argumentation theory. They wonder whether or not exceptional

circumstances are taken into account before generalizing - for instance in the case of the Shell managing director on the basis of which the conclusion is drawn that choosing an exact science as a subject will bring about a high salary. It appears that what laymen are wondering about comes close to Walton's (1996) question: "Were there special circumstances present in the example that would impair its generalizability?" (p. 50), for which I suggested that it might in fact be a heuristic question to answer the typicality-question (see 4.2.2). I also stated that in argumentation theory, it is usually unclear what the typicality/representativeness questions are aimed at; perhaps it is more sensible to replace these by questions explicitly asking for atypical or exceptional circumstances, for instance: 'is the generalization based on examples that may involve atypical circumstances?'

Second, laymen focus on the number of examples as suggested in argumentation theory. That is interesting, because it is also noted in argumentation literature that the number of examples should not be a reason to reject the argument, as it is typical that only a few examples are given. It appears from the closed interviews that if three individual cases are contrasted with one, number matters. Respondents find it important for several reasons, among which the diversity of characteristics that can be established by presenting multiple cases. As I already explained in 4.2, the diversity principle is common in qualitative research practice, so in that respect, it is remarkable that normative argumentation theory has not formulated a requirement for diversity (except for Windes & Hastings, 1969). So, if multiple cases are cited, the following question should be asked: 'Are the characteristics in the examples diverse?'

Third, a heuristic question could be asked to help determine the relevance of the argument: 'Are instances in the argument (clear) members of the conclusion categories?'. In addition, for the argument from example to a causal generalization, a distinction can be made between relevance of the action in the argument and relevance of the effect in the argument.

Last, it can be argued that in argumentation theory, a distinction should be made between reasoning from example to a *descriptive* generalization and reasoning from example to a *causal* generalization as Windes and Hastings (1969) do, because causality plays a role in laypersons' evaluation of the latter mode of reasoning and should, in my opinion, play a role in evaluation. Then, within reasoning from example to a causal generalization, another refinement is possible: one type of argument in which reasoning goes

from *causal* examples to a causal generalization and one type of argument in which reasoning goes from *correlational* examples to a causal generalization. In this study, my aim was to focus on the former type.

There are some limitations to this study. The design of the open cases in this study did not evoke the intended responses. The intention was for the respondents to spontaneously generate one strong example and one weak example in support of the causal generalization to find out what properties make a good example. It turned out that instead, respondents tended to present a statistical kind of generalization as a strong argument and a single example as a weak argument, or even other types of argument. The task or the task instruction might be regarded as unsuccessful in this respect (if the respondents were asked to read their arguments aloud and it became clear that they had not given the 'proper' arguments, they were not corrected). Hence, this raises the question as to what kind of criteria would have been generated if respondents had compared a single strong example with a single weak example.

Second, in the open case, the respondents filled in the following sentence: "Because I know a case in which that held true:...". That means that the respondents were not given the opportunity to give more than one individual example and as a result, the number-of-examples-criterion was not to be expected in the open case interviews. Nevertheless, some respondents referred to this particular criterion in the open case interviews as they compared a group of cases with a single case after all, but this appeared to be rather coincidental.

Third, although I aimed at the argument from *causal* examples to a causal generalization as I mentioned above and not the argument from *correlational* examples to a causal generalization, the closed case material appeared to have contained both types. In some arguments, causality was expressed, e.g. "Robin van Persie eats a balanced diet and is therefore in good condition". In other arguments, correlation was expressed, e.g. "Robert de Graaf started drinking more water and his skin looks less sallow now". In other words, the argument type I was interested in was not adequately operationalized.

All in all, this study revealed that there is little discrepancy between theory and laymen when it comes to the quality of the argument from example; most criteria laymen use do not conflict with argumentation theory. The number of examples criterion, the example accuracy criterion, the relevance criterion and the causality criterion

are shared by a considerable number of laymen in this study. Therefore, these criteria are likely to be part of laymen's argumentative baggage. In addition, criteria concerning causality and the relevance of action and effect are specific to the argument from example to a *causal* generalization, the subtype of the argument from example investigated in this study.

**PART II: THE RELATION BETWEEN ARGUMENT  
QUALITY AND ACTUAL PERSUASIVENESS:  
AN EXPERIMENTAL APPROACH**



## 5. The perceived reasonableness and actual persuasiveness of normatively strong and normatively weak arguments<sup>92</sup>

### 5.1 Introduction

In part I of this dissertation, I addressed the question to what extent laymen criteria for the argument from authority, the argument from cause to effect and the argument from example correspond to those formulated in argumentation theory. In chapter 1, it was argued that according to the ELM, normatively strong arguments should be more persuasive than normatively weak arguments in conditions in which messages are processed centrally. However, according to empirical findings, there may be a difference between what people believe to be a strong argument and what actually is persuasive (e.g. Hoeken, 2001a). What people believe to be a strong argument is expressed in the evaluation criteria they use to distinguish strong arguments from weak arguments. However, arguments that do conform to those criteria, or *normatively strong arguments*, are not necessarily more persuasive than arguments that violate those criteria, or *normatively weak arguments*. Therefore, the following research question was formulated:

Are normatively strong arguments from authority, arguments from cause to effect and arguments from example actually more persuasive than normatively weak arguments?

This will be the central research question in part II of the dissertation. An experiment was conducted to address this question. This experiment will be reported in chapter 6. However, before this experiment was conducted, a preliminary study was carried out. This study was performed in order to determine if it would be wise to measure other dependent variables besides actual persuasiveness. After all, the assumption underlying the second research question is that normatively strong arguments will be actually more persuasive because they are conceived as stronger arguments. Therefore, to test this assumption, it appears wise to first explore the relation between

---

<sup>92</sup> This study has been published earlier (Timmers, Šorm, Schellens & Hoeken, 2008). This publication is the result of an intensive collaboration between the first and the second author who contributed in equal measure to the design, execution, and reporting of the study.



the concept of what I call *perceived reasonableness* and the concept of *actual persuasiveness*.

In this chapter, a preliminary study will be reported in which the relation between perceived reasonableness and actual persuasiveness was investigated. First, I will further elaborate on the concepts of perceived reasonableness and actual persuasiveness and discuss if these concepts have already been investigated in previous empirical studies and if so, how they have been defined and operationalized.

### **Perceived reasonableness and actual persuasiveness**

In the study of argumentation, the concept of reasonableness is used to indicate the quality of arguments (Schellens, 1985; Van Eemeren, Garsen & Meuffels, 2007). Schellens (1985, p. 6) considers an argument reasonable if it, after being critically evaluated, increases the plausibility or acceptability of another statement. Empirical studies on the reasonableness of arguments are limited to fallacies, which are usually defined as unreasonable moves in argumentative discourse (Van Eemeren, 2001). Van Eemeren et al. (2007) carried out an extensive research project on ordinary language users' conceptions of reasonableness. They conducted a series of experiments, in which they investigated the perceived reasonableness of 24 different fallacies. In these studies, the concept of reasonableness has been consistently operationalized by asking participants to indicate on a seven-point scale how reasonable they found a discussion move. The results show that participants consistently consider fallacies to be less reasonable than their non-fallacious counterparts. Thus, the normative quality of an argument appears to influence the judgment of the reasonableness of that argument.

I wish to make three remarks on the research of Van Eemeren et al. (2007) that make additional research worthwhile. First, as the researchers were interested in conceptions of reasonableness, the actual persuasiveness has not been considered in these studies. In other words, it has not been tested whether fallacies that were found to be less reasonable than their non-fallacious counterparts were also less persuasive.

Second, to test the conventional validity of the argumentation scheme rule, the fallacies *argumentum ad consequentiam*, *argumentum ad populum*, *slippery slope* and *false analogy* were investigated. It is possible that ordinary judges are more sensitive to

the unreasonableness of such traditional fallacies than to the unreasonableness of normatively weak arguments that cannot be considered traditional fallacies. Moreover, as far as I know, no fallacy was used in the studies of Van Eemeren et al. (2007) that is related to one of the argument types that this dissertation focuses on.

Other empirical studies are concerned with the relation between what is *perceived* as persuasive and what actually *is* persuasive (e.g. Hoeken, 2001a; Dillard, Weber & Vail, 2007a; Dillard, Shen & Vail, 2007b). The studies of Dillard et al. (2007a; 2007b) show a substantial association between perceived persuasiveness and actual persuasiveness. Hoeken (2001a), by contrast, finds a difference between what people consider persuasive arguments and what they are in fact persuaded by.

The construct that is referred to in the study of Dillard, Weber and Vail (2007a), perceived persuasiveness, is apparently different from the concept of perceived reasonableness that I am interested in. Consider, for instance, how they defined the concept of perceived persuasiveness – for which they in fact use the term perceived effectiveness (PE):

“we inferred a conceptual definition of PE as an estimate of the degree to which a persuasive message will be favorably evaluated—in terms of its persuasive potential—by recipients of that message” (p. 617)

Hoeken (2001a) refers to the concept of perceived persuasiveness as “perception of the argument’s quality” (p. 429). It remains somewhat unclear whether that is the quality to persuade message recipients or the quality to be a reasonable argument. This ambiguity is reflected by the indicators that Hoeken (2001a) uses for perceived persuasiveness: “sound”, “relevant”, “strong” and “convincing”. Only the last indicator “convincing” clearly points at the intended concept of perceived persuasiveness. The indicators “sound” and “relevant” are more connected with perceived reasonableness. The indicator “strong” is an ambiguous term that may apply to perceived reasonableness as well as to perceived persuasiveness.

Dillard et al. (2007b) use the word pairs “convincing/not convincing”, “believable/not believable”, “sensible/not sensible”, “wise/foolish”, “right/wrong” and “important/unimportant” to measure perceived persuasiveness. For this study too, it goes that some indicators are more appropriate to measure perceived reasonableness than to measure perceived persuasiveness, namely “wise/foolish”, “sensible/not sensible” and “right/wrong”.

All in all, some scales in the studies of Dillard et al. (2007b) and Hoeken (2001a) unintentionally cover the concept of perceived reasonableness instead of the intended concept of perceived persuasiveness. This leads to the question of what the relation is between perceived reasonableness and actual persuasiveness, if perceived reasonableness is not (unintentionally) operationalized as a sub-aspect of perceived persuasiveness, but as a separate concept. It is possible that people will react differently to the question of what would persuade (perceived persuasiveness) than to the question of what would increase the plausibility or acceptability of another statement (perceived reasonableness). So, the association between perceived reasonableness and actual persuasiveness might be less substantial than the association between perceived persuasiveness and actual persuasiveness.

Furthermore, asking people to judge the reasonableness of an argument may stimulate them to focus on the supporting qualities of the argument. In judgments of claim acceptability, however, other factors besides supporting qualities might come to play a role, such as the opinions that people already hold. This means that someone may find a certain argument reasonable, but may nevertheless reject the claim that is supported by that argument. Therefore, it may be expected that the manipulation of argument quality by means of evaluation criteria will have more effect on perceived reasonableness scores than on claim acceptability scores.

Next, I will report a preliminary study on this issue. The study investigated the relation between perceived reasonableness and actual persuasiveness. The argument types that were selected to address this question were the argument from authority and the argument from cause to effect. This leads to the following research questions:

- (1) What is the relation between the perceived reasonableness and the actual persuasiveness of normatively strong and weak arguments from authority and normatively strong and weak arguments from cause to effect?
  - (1a) Are normatively strong arguments from authority perceived as more reasonable and are they more persuasive than normatively weak arguments from authority?
  - (1b) Are normatively strong arguments from cause to effect perceived as more reasonable and are they more persuasive than normatively weak arguments from cause to effect?

These questions were addressed in an experimental study. The method used in the experiment will be discussed in the next section.

## 5.2 Method

### 5.2.1 Material

The material consisted of 16 short texts in which the argument from authority or the argument from cause to effect was used. Each text was composed according to one of the following argument schemes:

Scheme (1) - argument from authority:

Source A says that measure M leads to effect E  
Therefore: measure M leads to effect E

Scheme (2) - argument from authority:

Source A says that effect E is desirable  
Therefore: effect E is desirable

Scheme (3) - argument from cause to effect:

Measures like M generally lead to effects like E  
Therefore: measure M will lead to effect E

Scheme (4) - argument from cause effect:

Measure M leads to effect G  
G leads to effect E  
Therefore: measure M will lead to effect E

In scheme (1), the argumentation results in a descriptive claim stating that a certain measure (an action or a policy) leads to a certain effect. In scheme (2), in which the argumentation results in a normative claim, the conclusion is drawn that the effect of a certain measure is desirable. In scheme (3), the conclusion is based on a general prediction stating that measures like the measure mentioned in the conclusion generally lead to effects like the effect mentioned in the conclusion. In scheme (4), the conclusion is based on two predictions, leading in two steps from the proposed measure to the proposed effect. The texts were composed in accordance with Table 5.1.

Table 5.1 Design with Argument Type, Argument Scheme, Criterion and Quality as independent variables.

	Argument from authority				Argument from cause to effect			
	Scheme (1)		Scheme (2)		Scheme (3)		Scheme (4)	
	Relevant Expertise	Trust-worthiness	Relevant Expertise	Trust-worthiness	Cause relevance	Effect relevance	Cause relevance	Effect relevance
Strong	(1)	(3)	(5)	(7)	(9)	(11)	(13)	(15)
Weak	(2)	(4)	(6)	(8)	(10)	(12)	(14)	(16)

Of the 16 texts, 8 were based on the argument from authority. Of these 8, 4 were composed according to scheme (1) and 4 according to scheme (2). Eight texts contained an argument from cause to effect. Of these 8, 4 used scheme (3) and 4 scheme (4). The result was that 4 claims expressed the desirability of a measure's effect, for instance: "It is a positive trend that more and more people practice two sports". Twelve other claims expressed an effect's occurrence as a result of a proposed measure, for instance: "Taking a cold shower after a tough game will lead to relaxation".

A pre-test was executed to be sure that the claims were not too acceptable or too unacceptable. In reasoning research, evidence has been found for the *belief bias effect*, which means that people are more tempted to accept a conclusion or argument when it corresponds with their beliefs than when it does not, regardless of the strength of the argumentation (see, e.g., Thompson, Striemer, Reikoff, Gunter & Campbell, 2003). A possible explanation for this effect is that reasoners are more likely to search for reasons to refute or disconfirm a conclusion they do not believe compared to a conclusion they find acceptable. In addition, Social Judgment Theory (Sherif & Hovland, 1961; Sherif, Sherif & Nebargall, 1965) predicts that if someone judges a message to lie within his latitude of rejection, attitude change will be reduced or nonexistent. In fact, the discrepant message may actually increase someone's position on the issue (*boomerang effect*) (Littejohn, 2002, p. 131). As prior beliefs may influence the acceptance of a conclusion or argument and as attitude change is less likely with messages laying within the latitude of rejection, I preferred to use claims in the experiment that were neither extremely accepted nor extremely unaccepted.

Fourty participants were asked to indicate for each of 32 claims on a seven-point scale to what degree they agreed with the claim (1 = strongly disagree; 7 = strongly agree). Claims scoring on

average between 3.00 and 5.00 were considered appropriate for the main experiment. Of the 32 pre-tested claims, 16 were selected.

The quality of the arguments was manipulated, resulting in a strong and a weak version of an argument in support of a claim. In the weak version, a criterion was violated, whereas in the strong version, that same criterion was met. The following criteria were selected for manipulating the argument from authority (scheme 1 and 2):

- (1) The relevant expertise criterion: the source's expertise is relevant to the opinion.
- (2) The trustworthiness criterion: the source should not have a vested interest in acceptance of the opinion.

The criteria that were selected for the manipulation of the argument from cause to effect (scheme 3 and 4) are:

- (3) The cause relevance criterion: the cause in the argument should connect adequately to the cause in the claim.
- (4) The effect relevance criterion: the effect in the argument should connect adequately to the effect in the claim.

These criteria were selected for several reasons: they have been formulated in argumentation theory, they are used by laymen, they can be considered specific to the argument from authority and the argument from cause to effect, and they also appeared appropriate for manipulation of argument quality (for further justification, I refer to section 6.1 in the next chapter).

Tables 5.2-5.5 contain a number of examples, showing how the arguments were composed according to a certain argument scheme and how the quality of the arguments was manipulated. All arguments in the examples are translated from Dutch.

Table 5.2 Example of the manipulation of the argument from authority according to scheme (1) and criterion (1).

Scheme (1) - the argument from authority	
Strong: Meets relevant expertise criterion	Weak: Violates relevant expertise criterion
Frans de Heer, who has been a club owner for over twenty years, states that the 'no alcohol' rule for young people under sixteen has led to a decline in street vandalism. Therefore, banning the use of alcohol among young people under sixteen will lead to a reduction in vandalism.	Erben Wennemars, first-class ice skater, states that the 'no alcohol' rule for young people under sixteen has led to a decline in street vandalism. Therefore, banning the use of alcohol for young people under sixteen will lead to a reduction in vandalism.

Table 5.3 Example of the manipulation of the argument from authority according to scheme (2) and criterion (2).

Scheme (2) - the argument from authority	
Strong: Meets trustworthiness criterion	Weak: Violates trustworthiness criterion
According to Karl Nooten, sports physiotherapist and fitness guru, it is a good development that more and more people do two sports. Therefore, it is a positive trend that more and more people practice two sports.	According to Annemiek van Rijn, working at the marketing department of the fitness club chain Health Groups, it is a good development that more and more people do two sports. Therefore, it is a positive trend that more and more people practice two sports.

Table 5.4 Example of the manipulation of the argument from cause to effect according to scheme (3) and criterion (3).

Scheme (3) - the argument from cause to effect	
Strong: Meets cause relevance criterion	Weak: Violates cause relevance criterion
Each form of publicity improves the reputation of a product. Therefore, an ad in a school paper will lead to a product become more known among young people.	Ads in the daily paper improve the reputation of a product. Therefore, an ad in a school paper will lead to a product become more known among young people.

Table 5.5 Example of the manipulation of the argument from cause to effect according to scheme (4) and criterion (4).

Scheme (4) - the argument from cause to effect	
Strong: Meets effect relevance criterion	Weak: Violates effect relevance criterion
A developing child who has a pet learns to treat the animal in a social and careful way. When a child is able to treat an animal in a social and careful way, it will be able to apply these skills to contact with people as well. Therefore, developing children who have a pet will have a better social interaction with other people than developing children without a pet.	A developing child who has a pet learns to take responsibility for taking care of the pet. If a child is able to take care of an animal, it will also be able to take care of people. Therefore, developing children who have a pet will have a better social interaction with other people than developing children without a pet.

Before the main experiment, a manipulation check was carried out to test if the manipulation of argument quality was perceived as such. This pre-test was designed as a forced choice test. Each participant received sixteen claims and in addition to each claim, the strong as well as the weak version of the supporting argument. Then participants had to indicate for each argument pair (A and B) whether or not they regarded one argument stronger than the other. An

example of such an item can be found in Table 5.6. The exemplary item is translated from Dutch.

Table 5.6 Exemplary manipulation check item<sup>93</sup>.

Therefore, immigration of highly educated young people to the Netherlands is necessary.	
A	B
There is an urgent need for a substantial immigration of highly educated young people from European and non-European countries to the Netherlands, says Dr. Karel van Lier of the Social and Cultural Planning Office in his report on the aging population problem.	There is an urgent need for a substantial immigration of highly educated young people from European and non-European countries to the Netherlands, says Dutch teacher Karel van Lier in a letter to the Volkskrant editor on the aging population problem.
Choose from the following options: <input type="radio"/> I find A strongest in supporting the claim. <input type="radio"/> I find B strongest in supporting the claim. <input type="radio"/> I find A and B equally strong in supporting the claim.	

Just as in the main experiment, eight claims were supported by arguments from authority according to scheme (1) and (2); eight claims were supported by arguments from cause to effect according to scheme (3) and (4). Four different versions of the questionnaires were created by varying the argument types (authority/cause to effect) and varying the position of the strong or weak argument (left/right).

25 participants took part in the forced choice test, 13 men and 12 women. The ages varied from 21 to 63. The average age was 42.4 ( $SD = 14.36$ ). The level of education varied from primary school to university education. All participants were laymen in the field of argumentation studies.

For each claim, it was tested by means of a chi-square test whether the strong argument or the weak argument was preferred. The results show that except for one item, there was a preference for the strong versions of the arguments in support of the claims ( $\chi^2(2)$ -values  $> 6.31$ ,  $\chi^2(1)$ -values  $> 11.56$ ,  $p$ -values  $< .05$ )<sup>94</sup>. The

<sup>93</sup> This item shows the original layout of the manipulation check items. The claim was presented above the alternative arguments. This layout does not correspond to the target items, in which the claim was presented after the argument.

<sup>94</sup> One item did not show a significant effect ( $\chi^2(1) = 1.00$ ,  $p = .32$ ). However, we can maintain that this item has been successfully manipulated, as none of the respondents preferred the weak version of the argument. Furthermore, the reason for degrees of freedom (df) varying, is that for some items, not all values occurred in the dataset. More specifically: three different values could occur



manipulation of the arguments in the strong and weak versions appears to be successful.

### 5.2.2 Instrumentation

Sixteen claims and supporting arguments were presented in a questionnaire titled "Opinions on social issues". The questionnaire started with an instruction, in which an example was given. In addition, it emphasized that no wrong answers could be given. The argument was always presented before the conclusion. For example:

In winter, light therapy causes people with stressful jobs to be less stressed. So, if you are suffering from a small winter depression, a few treatments with infrared light will improve your mood.

Two different kinds of questionnaires were used: one aimed at perceived reasonableness and the other aimed at actual persuasiveness. In the questionnaire about perceived reasonableness, the respondents were asked to indicate for each item on a seven-point scale to what degree they agreed with the statement that the argument is a good support for the conclusion in the last sentence (1 = strongly disagree; 7 = strongly agree). In the questionnaire about actual persuasiveness, the respondents were asked to indicate for each item on a seven-point scale to what degree they agreed with the last sentence (1 = strongly disagree; 7 = strongly agree).

Each dependent variable was measured with one single scale instead of multiple scales. It has been argued regularly that multiple item-measures are better than single-item measures (for a recent review of arguments for and against multiple item-measures, see Bergkvist & Rossiter, 2007). However, it has also been argued that multiple items are unnecessary if the object being rated is concrete singular and the attribute (dimension of judgment) is concrete (Bergkvist & Rossiter, 2007; Rossiter, 2002). In this study there was only one object to be rated (namely the argument or last sentence in the text presented) and the assumption was that (nearly) all raters would describe 'quality of support' and 'agreement with the last sentence' identically (indicating the concreteness of the attributes).

---

(preference for strong; preference for weak; not any preference) and sometimes only two of those three values were given to an item, leading to only one degree of freedom.

Therefore, it was decided that using one item for each dependent variable would be sufficient (cf. Van Eemeren, Garssen & Meuffels, 2007, p. 261, defending the use of one item for the reasonableness concept).

At the end of each questionnaire, the respondent was asked to indicate age, sex and education.

### **5.2.3 Design**

Two questionnaires were used in the experiment. Each questionnaire was presented to a different sample. The two questionnaires contained the same sixteen items, but differed with respect to the dependent variable. The first questionnaire focused on the perceived reasonableness of strong and weak arguments; the second was aimed at the actual persuasiveness of strong and weak arguments.

Each respondent judged either the strong or the weak version of the argumentation and in total judged eight strong and eight weak arguments. The order of the items was varied by rotating argument type (authority/cause to effect) and argument quality (strong/weak). This resulted in four different versions of a questionnaire.

### **5.2.4 Respondents**

The two samples consisted of 100 respondents, recruited by students in Dutch Language and Literature among their acquaintances, friends and family members. Half of them filled out the perceived reasonableness questionnaire and the other half filled out the actual persuasiveness questionnaire. The questionnaires were randomly assigned to the respondents. The male-female ratio in both groups was 50:50 and 48:52, respectively. In the perceived reasonableness group, the age varied between 17 and 58, with an average age of 36 ( $SD = 14.33$ ). In the actual persuasiveness group, the age of the respondents varied between 18 and 59, with an average age of 33 ( $SD = 13.18$ ). The level of education varied from primary school to university education. All respondents were laymen in the field of argumentation studies.

### **5.2.5 Procedure**

The respondents were asked to participate voluntarily in a study, in which their opinions on different social issues would be questioned. If they agreed, they received a questionnaire and any questions were

answered. The actual purpose of the study was revealed after they had completed the questionnaire. Completing the questionnaire took approximately fifteen minutes.

### **5.3 Results**

A multivariate analysis of variance with repeated measures (MANOVA) was used to investigate for each argument type the effects of argument quality ('Quality'), argument scheme ('Scheme') and criterion used for manipulation ('Criterion') on perceived reasonableness and actual persuasiveness ('Type of Judgment'). The decision was made to carry out the MANOVA's for each argument type, because the variables Scheme and Criterion for the argument from authority differed from the variables Scheme and Criterion for the argument from cause to effect.

#### **5.3.1 Argument from authority**

First, the question was addressed what the relation was between the perceived reasonableness and the actual persuasiveness of normatively strong and weak arguments from authority (research question 1). In other words, it was determined if the difference between normatively strong and normatively weak arguments was influenced by Type of Judgment (perceived reasonableness and actual persuasiveness). If this were the case, then interaction would occur between Type of Judgment (perceived reasonableness and actual persuasiveness) and Quality. The results, however, show that this effect did not occur ( $F < 1$ ). A main effect of Type of Judgment did not appear either ( $F < 1$ ).

From the analyses of the between subjects factor Type of Judgment, it appears that this factor does not have any influence. As a consequence, one single answer will be given for research question 1a: are normatively strong arguments from authority perceived as more reasonable and are they more persuasive than normatively weak arguments from authority? In Table 5.7, the results are presented for perceived reasonableness and actual persuasiveness.

Table 5.7 Perceived reasonableness and actual persuasiveness of the argument from authority in function of argument quality, argument scheme and criterion (*SD* between parentheses).

Argument scheme and criterion		Perceived Reasonableness		Actual Persuasiveness	
		Strong	Weak	Strong	Weak
Scheme 1:	Source A says that measure M leads to effect E Therefore: measure M leads to effect E				
	Relevant expertise criterion	4.04 (1.80)	3.58 (1.67)	4.08 (1.74)	3.76 (1.62)
	Trustworthiness criterion	4.56 (1.99)	3.34 (1.91)	4.58 (1.49)	3.36 (1.61)
Scheme 2:	Source A says that effect E is desirable Therefore: effect E is desirable				
	Relevant expertise criterion	3.74 (1.95)	2.84 (1.56)	3.76 (1.56)	2.88 (1.49)
	Trustworthiness criterion	3.84 (1.86)	3.34 (1.48)	4.24 (1.57)	3.74 (1.56)

The results showed significant main effects of Quality ( $F(1,98) = 31.11, p < .001, \eta^2 = .24$ ), Scheme ( $F(1,98) = 13.55, p < .001, \eta^2 = .12$ ) and Criterion ( $F(1,98) = 7.79, p < .01, \eta^2 = .07$ ). The interaction between Scheme \* Criterion almost reached the required level of significance ( $F(1,98) = 3.61, p = .06$ ). The main effects were qualified by a significant threeway-interaction between Scheme \* Criterion \* Quality ( $F(1,98) = 6.44, p < .05, \eta^2 = .06$ ). There were no other interactions between the variables ( $F(1,98) < 1.26, p > .26$ ). The results indicate that differences between normatively strong and weak arguments from authority are dependent on the argument scheme used and the criterion used for manipulation.

Analyses were performed separately for each argument scheme. For scheme 1, the interaction between Criterion \* Quality reached the required level of significance ( $F(1,98) = 5.05, p < .05, \eta^2 = .05$ ). From a paired comparison analysis using the LSD test, no significant difference in judgment appeared between strong arguments meeting the relevant expertise criterion and weak arguments violating the relevant expertise criterion ( $p = .12$ ). However, strong arguments meeting the trustworthiness criterion led to more positive judgments than weak arguments violating the trustworthiness criterion ( $p < .001$ ).

For scheme 2, the interaction between Criterion \* Quality did not reach the required level of significance ( $F(1,98) = 1.94, p = .17$ ). The only significant effect that was found, was that strong arguments led to more positive judgments than weak arguments ( $F(1,98) = 21.04, p < .001, \eta^2 = .18$ ).

So, when defending a descriptive claim, violating the criterion of trustworthiness makes a difference in judgment, whereas violating the criterion of relevant expertise does not. When defending a normative claim, strong arguments are more positively judged than weak arguments, but that effect does not depend on the criterion that is used for manipulation.

### **5.3.2 Argument from cause to effect**

First, the question was addressed as to what the relation is between the perceived reasonableness and the actual persuasiveness of normatively strong and weak arguments from cause to effect (research question 1). So, it was determined if the difference between normatively strong and normatively weak arguments was influenced by Type of Judgment (perceived reasonableness and actual persuasiveness). It goes for this argument type as well that an interaction should occur between Type of Judgment and Quality. The results show that this interaction nearly reached the conventional levels of significance: ( $F(1,98) = 3.54, p = .06$ ). The mean scores suggest that the judgments on the reasonableness of arguments from cause to effect are more extreme than the judgments on the acceptability of the conclusion of the same arguments. Hence, a weak argument from cause to effect is found to be weaker when one has to judge reasonableness than when one has to judge the acceptability of the conclusion. However, this interpretation should be dealt with carefully, as the effect did not meet conventional levels of significance. A main effect of Type of Judgment did not occur ( $F < 1$ ).

From the analyses of the between-subjects-factor Type of Judgment, it appears that this factor does not have a main effect. Therefore, one single answer will be given for the research question 1b: are normatively strong arguments from cause to effect perceived as more reasonable and are they more persuasive than normatively weak arguments from cause to effect? In Table 5.8, the results are presented for perceived reasonableness and actual persuasiveness.

Table 5.8 Perceived reasonableness and actual persuasiveness of the argument from cause to effect in function of argument quality, argument scheme and criterion (*SD* between parentheses).

Argument scheme and criterion		Perceived Reasonableness		Actual Persuasiveness	
		Strong	Weak	Strong	Weak
Scheme 3:	Measures like M generally lead to effects like E Therefore: measure M will lead to effect E				
	Cause relevance criterion	4.72 (1.65)	3.70 (1.76)	4.72 (1.57)	4.06 (1.65)
	Effect relevance criterion	4.32 (1.68)	3.92 (1.97)	3.92 (1.54)	3.58 (1.46)
Scheme 4:	Measure M leads to effect G G leads to effect E Therefore: measure M will lead to effect E				
	Cause relevance criterion	5.30 (1.37)	4.36 (1.89)	4.88 (1.22)	4.42 (1.39)
	Effect relevance criterion	4.30 (2.01)	3.00 (1.58)	3.98 (1.53)	3.74 (1.69)

The main effect of Quality is significant ( $F(1,98) = 26.48, p < .001, \eta^2 = .21$ ), just as the main effect of Criterion ( $F(1,98) = 43.98, p < .001, \eta^2 = .31$ ). The main effect of Criterion is qualified by a significant interaction between Scheme \* Criterion ( $F(1,98) = 7.67, p < .01, \eta^2 = .07$ ) and by a significant interaction between Scheme \* Criterion \* Type of Judgment ( $F(1,98) = 4.41, p < .05$ ). The main effect of Scheme was not significant ( $F(1,98) = 1.73, p = .19$ ). There were no other interactions between the variables ( $F(1,98) < 2.35, p > .12$ ). These results show that normatively strong arguments from cause to effect are perceived as more reasonable and are more persuasive than normatively weak arguments from cause to effect.

Analyses were performed separately for each type of judgment. For the perceived reasonableness judgments, the interaction between Scheme \* Criterion reached the required level of significance ( $F(1,49) = 11.40, p = .001, \eta^2 = .19$ ). For the actual persuasiveness judgments, the interaction between Scheme \* Criterion did not reach the required level of significance ( $F(1,49) < 1, p = .63$ ). This indicates that arguments manipulated on the basis of the cause relevance criterion lead to more positive judgments than

arguments manipulated on the basis of the effect relevance criterion, except when it concerns perceived reasonableness judgments of scheme 3 – the scheme based on a general causal relation.

#### **5.4 Conclusion and discussion**

The central research question was: what is the relation between the perceived reasonableness and the actual persuasiveness of normatively strong and weak arguments? We did not find significant differences between the perceived reasonableness and the actual persuasiveness of arguments. This result indicates that these two concepts are strongly related. Argumentation that is found to be more reasonable is generally also more persuasive in conditions in which central processing occurs. This is in line with Dillard et al. (2007a; 2007b) who showed a substantial association between perceived persuasiveness and actual persuasiveness.

The fact that no significant difference has been found between perceived reasonableness and actual persuasiveness may be the result of the claims that were used in this study. These claims were not too acceptable or too unacceptable. Had we used more 'extreme' claims, claims about which people were more likely to have an opinion, a significant difference might have been found. After all, the opinions that people already hold possibly influence their claim acceptability scores, so that the difference between claim acceptability scores and perceived reasonableness scores becomes more salient.

It should be noted that for the argument from cause to effect, the difference between the perceived reasonableness and the actual persuasiveness of normatively strong and weak arguments almost reached the conventional level of significance. Possibly, the difference between perceived reasonableness and actual persuasiveness of normatively strong and weak arguments is asymmetrical: if a claim is supported by an argument found reasonable by respondents, then they will be tempted to accept this claim. If a claim is supported by an argument that respondents find unreasonable, then they can nevertheless accept the claim, because they themselves have strong arguments in support of that claim. This suggests that one can expect stronger effects of argument quality on reasonableness judgments than on claim acceptability judgments.

The answers to the research questions, which ask whether normatively strong arguments are perceived as more reasonable than normatively weak arguments and whether they are more persuasive,

are affirmative. For the argument from cause to effect, it is clear that arguments meeting the normative criteria are seen as more reasonable and are more persuasive than arguments that do not meet these criteria. For the argument from authority, the matter is more complex. The results clearly show that arguments from authority that meet the criteria are perceived as more reasonable and are actually more persuasive than the arguments from authority that violate the criteria. This effect, however, is dependent on the scheme that is used and the criterion used for manipulation. If a descriptive claim is presented, meeting the trustworthiness criterion leads to more positive judgments than violating the trustworthiness criterion, whereas a manipulation on the basis of the relevant expertise criterion does not make any difference. If a normative claim is presented, strong arguments from authority lead to more positive judgments than weak arguments from authority; the effect of argument quality does not depend on criterion. In this study, it is difficult to interpret three-way-interactions between the variables, as each combination of variables (Scheme \* Criterion \* Quality) has been measured with only one item. Further research should demonstrate whether or not this interaction is too item-dependent.

It can be stated that the perceived reasonableness and actual persuasiveness of arguments are influenced by differences in argument quality. Furthermore, the argument scheme that is used and the criterion that is manipulated have an effect on perceived reasonableness and actual persuasiveness. This corresponds to the results of Hoeken and Hustinx (2009). They showed that people react differently to different kinds of evidence. Evidence can be defined as data (facts or opinions) presented as proof for an assertion (Reynolds & Reynolds, 2002). From the study reported here, it appears that we should not only differentiate between different types of data, but also that we should consider different argument schemes if we want to investigate the relation between argument quality and perceived reasonableness or actual persuasiveness.

This preliminary study was intended to investigate the relation between perceived reasonableness and actual persuasiveness. The results imply that measuring only actual persuasiveness as a dependent variable appears to be sufficient. After all, what is more persuasive is generally also perceived to be more reasonable.





## **6. The actual persuasiveness of normatively strong and weak arguments from authority, arguments from cause to effect and arguments from example**

### **6.1 Introduction**

According to the Elaboration Likelihood Model, taking the central route to persuasion should involve a critical evaluation of message content. What people do during such a critical evaluation is underspecified, but it has become clear that in critical evaluation of message content, argument quality should play a crucial role in the persuasiveness of the message. The ELM predicts that if people are highly motivated and able to process a message attentively, strong arguments will probably result in persuasion, whereas weak arguments will not. What it *is* exactly that makes arguments strong or weak, however, has been ill-defined. Instead, in previous empirical ELM-research, argument quality has been defined in terms of its observed persuasive effects rather than in an independent, conceptual meaningful way (e.g. Areni, 2003; Areni & Lutz, 1988; O'Keefe, 1995, 2002; Hornikx, 2005a; Van Enschot-Van Dijk, Hustinx & Hoeken, 2003). This problematic definition of argument quality has led to this dissertation's research aim to clarify what exactly makes arguments strong and what makes them weak. The focus is on the argument types that may be used to support the pragmatic argument's probability claim, expressing that a certain act may lead to a certain consequence. The argument types concerned are the argument from authority, the argument from cause to effect and the argument from example.

The previous studies reported in this dissertation have revealed the quality criteria for the argument from authority, the argument from cause to effect and the argument from example, as formulated in argumentation theory and according to laymen (chapter 2-4). According to the ELM, normatively strong arguments should be more persuasive than normatively weak arguments in conditions in which messages are processed centrally. There is a possibility, however, suggested by several studies (Hoeken, 2001a; O'Keefe, 1993), that arguments that do conform to criteria for argument quality are not necessarily more persuasive than those arguments that do not conform to criteria for argument quality. Therefore, the research question I intend to address in this chapter, is:

Are normatively strong arguments from authority, arguments from cause to effect and arguments from example actually more persuasive than normatively weak arguments?

The previous chapter (5) reported a study on the perceived reasonableness and actual persuasiveness of normatively strong and normatively weak arguments. The results showed that both variables are related - arguments that are perceived as more reasonable are generally more persuasive - and that they are influenced by differences in argument quality - normatively strong arguments are regarded as more reasonable and are more persuasive than normatively weak arguments. In the current study, only actual persuasiveness of the arguments will be investigated, not perceived reasonableness.

Although the current study is more limited with respect to the dependent variables, it is more elaborate than the previous one with respect to the independent variables. In the study reported in chapter 5, two argument types were investigated: the argument from authority and the argument from cause to effect. For the argument from authority, two criteria were used to manipulate normative argument quality: the relevant expertise criterion and the trustworthiness criterion. For the argument from cause to effect, the cause relevance criterion and the effect relevance criterion were used. The experimental study reported in this chapter is more elaborate than the one reported in chapter 5: the argument from example is added to the experimental design and more criteria are used for manipulating the quality of the argument from authority and the argument from cause to effect. Next, I will discuss the criteria that were chosen to investigate the impact of meeting and violating certain criteria on actual persuasiveness.

### **Selection of normative evaluation criteria**

To answer the research question above, a careful consideration should be made as to how to operationalize 'normatively strong arguments' and 'normatively weak arguments'. In other words, a careful decision must be made on which criteria to select for manipulating argument quality. In this study, the starting point was a list of *scheme-specific* criteria that were mentioned by argumentation experts *as well as* by laymen. To recapitulate the findings reported in chapter 2-4, the scheme-specific laymen criteria that are also mentioned in argumentation theory are the following:

For the argument from authority:

- The relevant expertise criterion: the source's expertise should be relevant to the opinion.
- The expert criterion: the source should be a genuine expert.
- The trustworthiness criterion: the source should be trustworthy.
- The external consistency criterion: the source's statement should be externally consistent.
- The internal consistency criterion: the source's statement should be internally consistent.
- The ability to provide evidence criterion: the source should be able to provide evidence.
- The recency criterion: the source's statement should be recent.

For the argument from cause to effect:

- The cause sufficiency criterion: the cause should be enough to probably bring about the effect.
- The cause relevance criterion: the cause in the argument should connect adequately to the cause in the conclusion.

For the argument from example:

- The number of examples criterion: there should be enough examples cited.
- The example accuracy criterion: the example should correspond to reality.
- The relevance criterion: the example should connect adequately to the conclusion.
- The typicality criterion: the example should be typical of the kinds of cases the generalization ranges over.
- The counterexample criterion: there should not be a counterexample undermining the conclusion.

Next, it was determined which of these criteria were suitable for the manipulation of argument quality. In Table 6.1, it is indicated for each of the criteria above whether or not it could be used for manipulation and if not, what the reason was.

Table 6.1 Suitability of criteria for manipulation of argument quality.

<i>Criteria</i>	<i>Suitable for manipulation?</i>
For the argument from authority:	
• The relevant expertise criterion	Yes.
• The expert criterion	Yes.
• The trustworthiness criterion	Yes.
• The external consistency criterion	No. Adding other sources with equal/other opinions to the argument changes the basic argument scheme, in which only one source is cited.
• The internal consistency criterion	No. Adding inconsistent opinions from the same source will create an unnatural and incredible text, as in

	'real' argumentation, we are rarely confronted with a supporting argument containing contradictory statements from the same source. In addition, adding opinions from the same source will change the basic argument scheme, in which one opinion is cited and not several.
• The ability to provide evidence criterion	No. Adding/removing evidence creates length differences at the same time. So, if an effect on actual persuasiveness would be found, an alternative explanation ('one text is longer than the other') may be valid.
• The recency criterion	Yes.
For the argument from cause to effect:	
• The cause sufficiency criterion	Yes.
• The cause relevance criterion	Yes.
For the argument from example:	
• The number of examples criterion	Yes.
• The example accuracy criterion	No. By making examples more or less accurate, the respondent's knowledge of the world is relied upon too much. It is dangerous to assume that someone knows whether a given example is true in real life or not.
• The relevance criterion	Yes.
• The typicality criterion	Yes.
• The counterexample criterion	No. It is impossible to manipulate an argument from example on the basis of this aspect. Whether or not there is an undermining counterexample is fully dependent on the respondent's knowledge or imagination.

After determining the suitability of each criterion for manipulation, a few other decisions were made. First, it was decided to add the following criteria to the study:

For the argument from authority:

- Opinion-to-conclusion relevance criterion: the source's opinion should be relevant to the conclusion.

For the argument from cause to effect:

- The effect relevance criterion: the effect in the argument should connect adequately to the effect in the conclusion.

For the argument from example:

- The diversity criterion: the examples should be diverse.

I used the criterion that the source's statement should be relevant to the claim at hand, as this criterion can be considered to be a specific interpretation of the general relevance criterion (see 2.2.2). After all, if the source's presumed opinion is insufficiently connected to the conclusion it is supposed to support, it is less relevant to the matter at hand. Thus, it is relevant from a theoretical perspective to investigate arguments from authority that do or do not meet this criterion. Furthermore, the criterion appeared suitable for manipulation.

Second, the criterion that the effect should be relevant, associated with the argument from cause to effect, was used in this study. This criterion was not mentioned in theory and it was observed in the empirical study that the respondents never focused on the relevance of the effect only. However, it was explained in chapter 3 that this criterion is reasonable from a theoretical perspective (see 3.2.3). Furthermore, the observation that laymen do not focus on the effect in the argument can be explained by overly subtle manipulations of the effects in the closed case (see 3.4). The criterion also appeared to be suitable for manipulation. Therefore, in this study, I decided to apply the relevance criterion to the effect as well.

Third, I used the criterion that examples should be diverse, which is associated with the argument from example. This criterion is generally not mentioned in argumentation theory (except for Windes & Hastings, 1969), but I stated earlier that this appears to be a reasonable criterion (see 4.2.3). Furthermore, it became clear from the empirical study on laymen criteria that laymen do pay attention to diversity (see 4.3). The diversity criterion can also be used for manipulation. Therefore, this criterion was selected for the current study.

Finally, to optimize the balance of conditions across the argument types, the decision was made to leave out the recency criterion, pertaining to the argument from authority<sup>95</sup>. For the same reason, I did not use the relevance criterion, pertaining to the argument from example.

In sum, the following criteria were used in the current study:

For the argument from authority:

- (1) The expert criterion
- (2) The relevant expertise criterion
- (3) The trustworthiness criterion
- (4) The opinion-to-conclusion relevance criterion

---

<sup>95</sup> I refer to Timmers (to appear), who did use the recency criterion in a similar experiment.

- For the argument from cause to effect:
- (5) The cause sufficiency criterion
  - (6) The cause relevance criterion
  - (7) The effect relevance criterion

- For the argument from example:
- (8) The typicality criterion
  - (9) The number of examples criterion
  - (10) The diversity criterion

A remark should be made with respect to the expert criterion (1) and the relevant expertise criterion (2), associated with the argument from authority. In the previous study, the relevant expertise criterion (the expertise of the source should be relevant to the field the statement is in) appeared to be mixed with the expert criterion (someone should have expertise, regardless of the field the statement is in). Here is an example to illustrate this problem (the example was also presented in chapter 5):

Strong: Meets relevant expertise criterion	Weak: Violates relevant expertise criterion
There is an urgent need for a substantial immigration of highly educated young people from European and non-European countries to the Netherlands, says Dr. Karel van Lier of the Social and Cultural Planning Office in his report on the aging population problem. Therefore, immigration of highly educated young people to the Netherlands is necessary.	There is an urgent need for a substantial immigration of highly educated young people from European and non-European countries to the Netherlands, says Dutch teacher Karel van Lier in a letter to the Volkskrant editor on the aging population problem. Therefore, immigration of highly educated young people to the Netherlands is necessary.

If the strong argument is more persuasive than the weak argument, it cannot be determined what exactly caused the effect: was it the fact that Dr. Karel van Lier in the strong argument has a higher degree than the teacher Karel van Lier in the weak argument? Or is the difference due to the fact that in the strong version, Karel van Lier's knowledge is apparently more relevant to the opinion than in the weak version? In other words: the property of being more credible as an expert source (expert criterion) was confounded by having relevant expertise (relevant expertise criterion). Therefore, in this study, I decided to disentangle these two properties of the argument from authority. This decision can also be justified by the fact that in theory and in practice, these two criteria also appear to be disconnected (see chapter 2).

An experiment was conducted to answer the research question above. For this experiment, a number of short texts were constructed. In each of these texts, either an argument from authority, an argument from cause to effect or an argument from example was used. Different variants of an argument were created: strong ones, meeting particular criteria, and weak ones, failing to meet these standards. The respondents were asked to judge the degree to which they (dis)agreed with the claim supported by the argument. In a pre-test, I collected agreement ratings on claims only, without any arguments. The purpose of this pre-test was to select relevant stimuli for the experiment. In the next section, the method of the study will be described.

## 6.2 Method

### 6.2.1 Pre-test

A pre-test was conducted to select claims that are not perceived as extremely acceptable or extremely unacceptable (for an explanation, see the previous chapter, section 5.2.1). A set of 70 claims was composed, including 5 claims that were expected to be perceived as extremely acceptable and another 5 claims that were expected to be perceived as extremely unacceptable. The 10 'extreme' claims can be considered as anchor points, giving respondents the opportunity to give extreme scores once in a while. These 10 'extreme' claims were included to prevent respondents exaggerating the differences between the 60 'moderate' claims and judge them as relatively acceptable or unacceptable. Each of the 70 claims, listed in appendix 6A, contained an act leading to a consequence. For instance: "Eating lots of vegetables leads to a healthier blood pressure". Two versions of the list were created to counter possible order effects<sup>96</sup>.

Respondents received one of the two versions, and were requested to rate each claim on a seven-point Likert scale (1 = Completely disagree; 7 = Completely agree) to what extent they

---

<sup>96</sup> Version 1: the order in which the claims were presented corresponds to the order in appendix 6A. Version 2: was version 1 in reverse order. I checked whether there were any order effects. More specifically, I checked whether the effect of claim on agreement depended on version. I conducted a repeated measures analysis of variance, where claim was a within-subjects factor and version a between-subjects factor. If version had influenced the relationship between claim and agreement, a significant interaction effect between claim and version should have occurred. However, this interaction effect did not appear:  $F(69, 2622) = 1.13, p = .216$ .



agreed with that claim. The pre-test was held in a public library in Nijmegen, the Netherlands. Forty people participated, of which 40% were male, 52.5% were female and 7.5% did not indicate their sex. The mean age of the respondents was 45.50 ( $SD = 14.81$ ) and ranged from 19 to 74. Their level of education ranged from primary education to university education, and most respondents had received higher vocational education (40%)<sup>97</sup>. Each respondent received 10 Euros for participating. The procedure took 10 to 15 minutes.

Appendix 6A lists the 70 claims and the mean agreement ratings on a seven-point scale. Claims for which the mean agreement ratings ranged from 3.00 to 5.00 were considered suitable for the experiment. From a pool of 41 suitable claims, 34 claims were selected for the experiment.

### **6.2.2 Material**

The material used in the experiment consisted of short texts. Each text consisted of a claim and a supporting argument. The claim and argument were always presented in the same order: first the argument, then the claim. For claims selected on the basis of the pre-test results, arguments were developed in accordance with Table 6.2.

---

<sup>97</sup> Three respondents did not fill out whether they were male or female, two respondents did not fill out their age and three respondents did not indicate their level of education.

Table 6.2 Experimental design.

Argument from authority					Argument from cause to effect					Argument from example				
Scheme (A):					Scheme (B):		Scheme (C):			Scheme (D):		Scheme (E):		
A says that M leads to E Therefore: M leads to E					M leads to G G leads to E Therefore: M leads to E		Measures like M generally lead to effects like E Therefore: M leads to E			In A1, M leads to E Therefore: M leads to E		In A1, M leads to E In A2, M leads to E Therefore: M leads to E		
Strong: meets criteria (1)-(4)	Weak: violates criterion (1)	Weak: violates criterion (2)	Weak: violates criterion (3)	Weak: violates criterion (4)	Strong: meets criterion (5)	Weak: violates criterion (5)	Strong: meets criteria (6) & (7)	Weak: violates criterion (6)	Weak: violates criterion (7)	Strong: meets criterion (8) violates criterion (9)	Weak: violates criteria (8) & (9)	Strong: meets criteria (8) & (9) violates criterion (10)	Strong: meets criterion (9) & (10) A2 violates criterion (8)	Weak: violates criterion (8) & (10) meets criterion (9)
(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)	(VIII)	(IX)	(X)	(XI)	(XII)	(XIII)	(XIV)	(XV)

Table 6.2 shows that the arguments varied in argument type, argument scheme and the criterion for argument quality that was violated. For the argument from authority, one argument scheme was used<sup>98</sup>. Four different evaluation criteria were applied to create the strong arguments and the weak arguments from authority. According to the criteria, a strong argument from authority is one in which the source is an expert (1) in a field relevant to the claim (2), who does not have a vested interest in the audience accepting this claim (3), and whose statement is relevant to the claim at hand (4). Apart from an authority who meets all these criteria, four versions were created. In each version, one of these criteria was violated whereas the others were met.

<sup>98</sup> Contrary to the design of the experiment reported in chapter 5, in which two different argumentation schemes were used - one with a descriptive claim and one with a normative claim - I only used a scheme with a descriptive claim. I refer to Timmers (to appear) who, in a similar experiment, focuses on the argument from authority supporting a normative claim.

Next, I present some examples. In the questionnaire, the texts were presented in Dutch. Here, the Dutch texts have been translated into English. I refer to appendix 6B for the Dutch source texts.

*Condition (I): argument from authority; scheme (A); strong: meets criteria (1)-(4).*

Dr. Masha van Rijn, who obtained her doctorate with a thesis on new media and business administration and currently works for the Department of Trade and Industry, claims that representatives show a higher involvement with their companies if they can use new media. Therefore: representatives feel more involved with their companies if they can use new media, such as a laptop with an internet connection, at home or on the way.

*Condition (II): argument from authority; scheme (A); weak: violates (1) the expert criterion.*

Masha van Rijn, a student participating in the HVE [Higher Vocational Education] programme New Media and Business Administration at the INHOLLAND University of Applied Sciences, claims that representatives show a higher involvement with their companies if they can use new media. Therefore: representatives feel more involved with their companies if they can use new media, such as a laptop with an internet connection, at home or on the way.

*Condition (III): argument from authority; scheme (A); weak: violates (2) the relevant expertise criterion.*

Dr. Masha van Rijn, who obtained her doctorate with a thesis on law and economics and currently works for the Department of Trade and Industry, claims that representatives show a higher involvement with their companies if they can use new media. Therefore: representatives feel more involved with their companies if they can use new media, such as a laptop with an internet connection, at home or on the way.

*Condition (IV): argument from authority; scheme (A); weak: violates (3) the trustworthiness criterion.*

Dr. Masha van Rijn, who obtained her doctorate with a thesis on new media and business administration and currently works for internet provider UPC, claims that representatives show a higher involvement with their companies if they can use new media. Therefore: representatives feel more involved with their companies if they can use new media, such as a laptop with an internet connection, at home or on the way.

*Condition (V): argument from authority; scheme (A); weak: violates (4) the opinion-to-conclusion relevance criterion.*

Dr. Masha van Rijn, who obtained her doctorate with a thesis on new media and business administration and currently works for the Department of Trade and Industry, claims that representatives are engaged in their

activities somewhat more intensively, if they can use new media. Therefore: representatives feel more involved with their companies if they can use new media, such as a laptop with an internet connection, at home or on the way.

For the argument from cause to effect, two different argumentation schemes were used: a scheme based on two causal connections and a scheme based on a causal generalization. For the former scheme, one criterion was used to develop the strong and the weak argument. According to this criterion, a strong argument from cause to effect is one in which the occurrence of the cause should probably bring about the effect (5) and a weak argument from cause to effect is one in which this criterion is violated. For the latter scheme, two criteria were used<sup>99</sup>. According to these criteria, a strong argument from cause to effect is one in which the cause in the conclusion is a clear case of the kind of causes in the argument (6) and in which the effect in the conclusion is a clear case of the kind of effects in the argument (7). Apart from an argument from cause to effect that meets these two criteria, two versions were created. In each version, one of these two criteria was violated whereas the other was met.

The following examples concern the argument from cause to effect (see appendix 6B for the Dutch source texts):

*Condition (VI): argument from cause to effect; scheme (B); strong: meets (5) the cause sufficiency criterion.*

Playing Mozart in home furnishing shops will have a positive effect on the mood of consumers. This will raise the turnover. So, playing Mozart's music in home furnishing shops leads to a higher turnover.

*Condition (VII): argument from cause to effect; scheme (B); weak: violates (5) the cause sufficiency criterion.*

Playing Mozart in home furnishing shops means consumers will stay there longer and listen to the music. This will raise the turnover. So, playing Mozart's music in home furnishing shops leads to a higher turnover.

*Condition (VIII): argument from cause to effect; scheme (C); strong: meets criteria (6) & (7).*

---

<sup>99</sup> In the experiment reported in chapter 5, these criteria were applied on both argumentation schemes. In the current experiment, the criteria were only used for the scheme based on a causal generalization. After closer inspection of the material in the previous experiment, it appeared more natural to apply the criteria on the scheme based on a causal generalization.

In most cases, home furnishing shops benefit financially if they play classical music. So, playing Mozart's music in home furnishing shops leads to a higher turnover.

*Condition (IX): argument from cause to effect; scheme (C); weak: violates (6) the cause relevance criterion.*

In most cases, home furnishing shops benefit financially if they play popular music. So, playing Mozart's music in home furnishing shops leads to a higher turnover.

*Condition (X): argument from cause to effect; scheme (C); weak: violates (7) the effect relevance criterion.*

In most cases, home furnishing shops benefit promotionally if they play classical music. So, playing Mozart's music in home furnishing shops leads to a higher turnover.

Two argumentation schemes were used concerning the argument from example: one based on a single example and the other based on two different examples. For the former scheme, a strong argument from example is one in which the example in the argument is typical of the category in the conclusion (8). In the weak version, this criterion was violated. The argumentation scheme that was based on a single example was inherently weak, as it automatically violated the number of examples criterion: there should be enough examples cited (9).

For the latter scheme, there was no absolute strong argument from example, because each version violated one or more criteria. However, two versions were stronger than a third version, as they only violated one criterion, whereas the third version violated two criteria. Therefore, I call these two versions *relatively* strong. A relatively strong argument from example is one in which both examples in the argument are typical of the category in the conclusion (8), but in which the examples are not diverse (10). Furthermore, a relatively strong argument from example is one in which the examples presented are diverse (10), but in which one of the examples is atypical (8). Apart from these two relatively strong versions, one relatively weak version was created that violated both criterion (8) and (10). In this version, both examples presented were atypical and hence not diverse. The argumentation scheme using two examples was inherently strong, as it automatically met the number of examples criterion: there should be enough examples cited (9). These are examples of texts containing an argument from example (see appendix 6B for the Dutch source texts):

*Condition (XI): argument from example; scheme (D); strong: meets (8) the typicality criterion; violates (9) the number of examples criterion.*

If people with a balance disorder play handball regularly, they can recover more quickly. Therefore, people with a balance disorder benefit from playing ball sports.

*Condition (XII): argument from example; scheme (D); weak: violates (8) the typicality criterion and (9) the number of examples criterion.*

If people with a balance disorder play water polo regularly, they can recover more quickly. Therefore, people with a balance disorder benefit from playing ball sports.

*Condition (XIII): argument from example; scheme (E); strong: meets (8) the typicality criterion and (9) the number of examples criterion; violates (10) the diversity criterion.*

If people with a balance disorder play handball regularly, they can recover more quickly. Or basketball: that also helps to develop balance. Therefore, people with a balance disorder benefit from playing ball sports.

*Condition (XIV): argument from example; scheme (E); strong: meets (9) the number of examples criterion and (10) the diversity criterion; A2 violates (8) the typicality criterion.*

If people with a balance disorder play handball regularly, they can recover more quickly. Or water polo: that also helps to develop balance. Therefore, people with a balance disorder benefit from playing ball sports.

*Condition (XV): argument from example; scheme (E); weak: A1 and A2 violate (8) the typicality criterion and (10) the diversity criterion; meets (9) the number of examples criterion.*

If people with a balance disorder play water polo regularly, they can recover more quickly. Or ice hockey: that also helps to develop balance. Therefore, people with a balance disorder benefit from playing ball sports.

This resulted in a design with a total of 15 conditions. Of each condition, there were two different cases. In other words: each condition was represented by two different texts. So, a questionnaire contained 10 texts with an argument from authority, 10 texts with an argument from cause to effect and 10 texts with an argument from example. A questionnaire consisted of a total number of 34 texts. The remaining four texts contained circular arguments<sup>100</sup>. These circular

---

<sup>100</sup> In this traditional fallacy, also known as *begging the question*, *petitio principii* or *circular reasoning*, "the arguer assumes that what needs to be proven (the question at issue) has already been shown to hold" (Van Eemeren, Grootendorst & Snoeck Henkemans, 1996, p. 68).

arguments were included in the material to be able to determine afterwards whether or not a respondent had studied the arguments attentively. After all, one might expect people who focus all their attention on the argumentation to recognize such traditional fallacies and not to be extremely persuaded by them. Here is an example of a text with a circular argument (see appendix 6B for the Dutch source text):

A government that keeps a better eye on clubs for adolescents prevents the youth from becoming mentally and physically dependent on beer. So if youth clubs are better supervised by the government, this will lead to less beer addiction among adolescents.

### 6.2.3 Respondents

The respondents were approached in (the vicinity of) the public library in Nijmegen, the Netherlands, in line with the pre-test of this study. No respondent had participated in any of the other studies reported in this dissertation. A total of 200 people responded to the paper-and-pencil questionnaire. Three questionnaires were excluded from the analyses because the scales were interpreted incorrectly and one because there were doubts about the respondent's mastery of the Dutch language. Furthermore, four respondents' data were rejected as they scored extremely high (seven on a seven-point Likert scale)<sup>101</sup> on all four claims supported by circular arguments. From these scores, it was inferred that these four respondents had not paid enough attention to the arguments. Finally, the data of five respondents were excluded from analysis, as they appeared to possess special knowledge in the field of argumentation<sup>102</sup>.

---

<sup>101</sup> To decide what would be an 'extremely high score', I took the data resulting from the pre-test as a starting point (see appendix 6A, nr. 14, 33, 35, 58). For each claim, I added up the value of one standard deviation to the mean agreement rating. Each outcome approached a value of six. So, I decided to consider a score of seven on the claims supported by circular arguments as an 'extremely high score'.

<sup>102</sup> To measure the degree of expertise in the field of argumentation, questions were asked about the concepts argument types, argument schemes and argument evaluation/tests (also see 'Instrumentation'). On each question, respondents could score zero (indicating "never heard of the concept"), one ("heard of the concept but does not give an example"), two ("heard of the concept but gives a bad example") and three ("heard of the concept and gives a good example"). Respondents who scored three on two out of three questions, were not considered laymen anymore.

On average, the respondents were 36.67 ( $SD = 17.65$ ) years old (range: 15 - 79)<sup>103</sup>. The percentage of male respondents was 39%; that of female respondents 61%. Their level of education varied from primary education to university education, but most respondents had received vocational education (27.3%) and university education (29.9%).

#### 6.2.4 Design

Ten versions of the questionnaire were made. A balanced Latin square design was used to establish the first five versions that varied in argument quality. For example, the first claim presented in version 1 was supported by a strong argument from authority; in version 2, the same claim was supported by an argument from an authority that was not credible as an expert source; in version 3, the same claim was supported by an argument from an authority that did not assert what was expressed in the claim; etcetera.

The next five versions were created by putting the texts in each of the first five versions in reverse order. For example, the first text in version 1 became last in version 6; the text on the second position became second last, etcetera. In this way, the position of the texts also varied across versions.

The texts with circular arguments were placed between the other texts. Their order of appearance was identical in the first five versions; in the next five versions, they appeared in reverse order.

#### 6.2.5 Instrumentation

The questionnaire was titled "Opinions on societal issues". Respondents were requested to indicate for each text on a seven-point Likert-scale the degree to which they (dis)agreed "with the conclusion in the last sentence". It was emphasized in the written instruction that it were their personal opinions that were relevant and that their answers could not be wrong. In addition, an example was presented.

After the 34 texts, control items were presented about the texts that each respondent had judged. The control items enabled a check whether or not the manipulations were successful. Each respondent received control items, but *which* items a respondent

---

<sup>103</sup> One respondent did not indicate his or her age.



received depended on the target items judged in the previous part of the questionnaire. For (a number of) target items that respondents had judged in the first part, corresponding control items were presented in the second part of the questionnaire.

First, control items were presented allowing a manipulation check of a part of the arguments from authority. In the control items, the same arguments from authority were presented as in the target items, but without the supported claim. For instance:

Dr. Masha van Rijn, who obtained her doctorate with a thesis on new media and business administration and currently works for the Department of Trade and Industry, claims that representatives show a higher involvement with their companies if they can use new media.

Respondents had to indicate on a seven-point Likert scale the degree to which “the underlined source is (in)competent in the field that the statement is in” (1 = very incompetent in the field that the statement is in; 7 = very competent in the field that the statement is in) and/or “the underlined source is (un)biased with respect to the field that the statement is in” (1 = very unbiased with respect to the field that the statement is in; 7 = very biased with respect to the field that the statement is in). Indeed, the source was underlined in the control items to avoid confusion among the respondents about what exactly was meant by “the source”.

Whether or not both scales were presented depended on the particular text the respondents had judged in the first part of the questionnaire. If they had judged a text representing condition I (meeting all criteria), both scales were presented; if they had judged a text representing condition III (violating the relevant expertise criterion), only the first scale was presented; if they had judged a text representing condition IV (violating the trustworthiness criterion), only the second scale was presented. The texts representing condition II (violating the expert criterion) and V (violating the opinion-to-conclusion relevance criterion) were not controlled for.

The reason for not controlling these manipulations in condition II and V was that in my opinion, the messages did or did not differ in source expertise (condition II) and relevance of the presumed opinion (condition V), independent of respondent perceptions. According to O’Keefe (2003), it is unnecessary to check the adequacy of the manipulation of such intrinsic measure properties (e.g. message length). After all, if the check fails, it cannot be concluded that the researcher must have manipulated the message property improperly.

No matter the respondent's perception of the property, the message *did* differ with respect to the property.

Second, control items were presented, allowing a check of the manipulation of arguments from cause to effect. More specifically, perceived probability was checked for condition VI (meeting the cause sufficiency criterion) and VII (violating the cause sufficiency criterion), in order to determine if it indeed differed significantly between conditions. For conditions VIII (meeting the cause relevance criterion and the effect relevance criterion), IX (violating the cause relevance criterion) and X (violating the effect relevance criterion), the perceived probability was also checked. However, here the objective was to determine if this factor could offer an alternative explanation for differences in persuasiveness, in case these were revealed.

In the control items, the same arguments from cause to effect were presented as in the target items, but without the supported claim. In addition, arguments representing conditions VI and VII were split up into two separate causal connections. For example, consider the following target item:

Playing Mozart in home furnishing shops means consumers will stay there longer and listen to the music. This will raise the turnover. So, playing Mozart's music in home furnishing shops leads to a higher turnover.

In the corresponding control item, the two causal connections that can be identified in the target argument were separately presented:

- (1) Playing Mozart in home furnishing shops means consumers will stay there longer and listen to the music.
- (2) If consumers in home furnishing shops stay longer and listen to the music, the turnover will be raised.

Arguments representing conditions VIII, IX and X were presented again in the control items. For instance:

In most cases, home furnishing shops benefit financially if they play classical music.

Respondents had to rate the causal connections in terms of (im)probability. They were asked to indicate for each sentence (containing a causal connection) on a seven-point Likert scale the degree to which they found "the relation between the cited events (im)probable" (1 = very improbable; 7 = very probable).

Third, control items were presented, allowing a check of the manipulation of arguments from example. The arguments from example were controlled for the typicality of the example. The respondents judged a proposition of the type '[member in argument] is a typical example of [*category in claim*]'. For instance, if someone had seen the target item

If people with a balance disorder play water polo regularly, they can recover more quickly. Therefore, people with a balance disorder benefit from playing ball sports.

the following corresponding check item was presented:

Water polo is a typical example of the category of *ball sports*.

Then respondents had to indicate on a seven-point Likert scale the degree to which they "(dis)agreed with the proposition" (1 = Completely disagree; 7 = Completely agree).

To avoid confusion among respondents, the control items concerning a particular argument type were grouped together. In each version, the control items regarding the argument from authority were presented first, then the control items with respect to the argument from cause to effect and last, the items for the argument from example. Within groups of control items, the order in which the items appeared varied across versions. As a result, for example, version 6 contained exactly the same control items as version 1, grouped the same way according to argument type, but the items within a group were presented in reversed order.

The questionnaire ended with questions about the respondents' age, sex, and level of education. In addition, the respondents were questioned about the perceived goal of the current study. Last, the respondents were to answer questions that enabled me to check the amount of knowledge in the field of argumentation. Specifically, the question was asked if respondents had come into contact with argumentation studies<sup>104</sup>, if they had ever heard of argument types, if they had ever heard of argument schemes, and finally, if they had ever heard of ways to evaluate or test arguments. If respondents answered with 'yes' in each of the last three cases, they were invited to give an example.

---

<sup>104</sup> I acknowledge that in English as well as in Dutch ('Bent u wel eens in aanraking gekomen met argumentatieleer?'), the formulation is awkward. A better alternative would have probably been if someone had ever been occupied with/engaged in argumentation studies.

### **6.2.6 Procedure**

Respondents willing to participate were brought into a separate room in the public library. Eight tables and chairs were placed in the room, so that groups of respondents could work simultaneously. As soon as the respondents had taken a seat, they were asked to read the instructions attentively and ask for clarification if necessary. After the questionnaire had been filled in, the respondents were rewarded with 10 Euros and thanked for their cooperation. The real research purpose was revealed, if asked for. The whole procedure took about 25 minutes.

### **6.2.7 Statistical tests**

The data on the argument from authority were analyzed by means of a 5 (condition) x 2 (case) repeated measures analysis of variance, where condition and case were both within-subject factors. As mentioned in the method section, each condition was represented by two different cases. It is possible that the persuasiveness differed between the first case and the second case judged by a respondent. After all, the second time the respondent was confronted with a certain condition, the particular manipulation may have been more likely to be noticed. Therefore, I involved case as a factor in the analysis. Given that I wanted to assess the impact of violating a criterion for argument quality, planned comparisons were conducted between the strong argument from authority on the one hand, and the other four conditions on the other hand (using the LSD test).

The data on the argument from cause to effect were analyzed for the two argument schemes separately. First I tested whether the strong argument using two causal steps was more persuasive than the weak argument using two causal steps. I used a 2 (condition) x 2 (case) repeated measures analysis of variance, where condition and case were both within-subject factors. Then, I tested the difference in persuasiveness for the three conditions in which causal generalizations were used. I used a 3 (condition) x 2 (case) repeated measures analysis of variance, where condition and case were both within-subject factors. I also performed planned comparisons (LSD), as I wanted to compare the strong variant with each of the two weak variants.

The data on the argument from example were analyzed by means of a 2 (number) x 2 (typicality) x 2 (case) repeated measures analysis of variance, where number, typicality and case were all

within-subject factors. In this particular analysis, I did not include the condition in which one typical and one atypical example was used. This condition, however, was included in a separate analysis I carried out for the three conditions in which two examples were used (scheme E), to compare the three conditions: two typical examples, one typical and one atypical example and two atypical examples. This was done by means of a 3 (condition) x 2 (case) repeated measures analysis of variance, where condition and case were within-subject factors. A planned comparison analysis (LSD) was conducted to compare the relatively strong versions with the relatively weak version.

## **6.3 Results**

### **6.3.1 Manipulation checks**

In this section, the results of the manipulation checks will be reported. I will indicate which manipulations were successful.

#### **The argument from authority**

I wanted to know whether the sources used in the strong arguments from authority were indeed perceived by the respondents as more competent in the field the statement is in ('relevant expertise') and less biased towards the field the statement is in ('bias') than the sources in the weak arguments from authority. More specifically, I expected the sources in the strong condition I to be perceived as more competent in the field than the sources in the weak condition III. I also expected the sources in the strong condition I to be perceived as less biased than the sources in the weak condition IV (see Table 6.2).

If the material had been manipulated successfully, independent-samples t-tests should have shown significant differences in scores for the conditions concerned. I used independent-samples t-tests, as I wanted to compare the mean scores in two different conditions: the strong condition I with the weak condition III and the strong condition I with the weak condition IV.

It appears that the majority of the material was successfully manipulated. There are two exceptions: for two texts (texts 5 and 6), the differences in mean scores between the two conditions I and III were not significant. That means that the sources in these text conditions were perceived as equally competent in the field the

statement is in (Appendix 6C1 shows the results of the manipulation checks. I refer to appendix 6B for texts 5 and 6.).

### **The argument from cause to effect**

Regarding the argument from cause to effect, I checked the perceived probability of the causal connections in each argument. For arguments using scheme (B), there were two causal connections to be judged. I compared the perceived probability scores for the *first* causal connection in the strong condition VI with the weak condition VII. I also compared the perceived probability scores for the *second* causal connection in the strong condition VI with the weak condition VII. For example, the scores for 'Playing Mozart in home furnishing shops will have a positive effect on the mood of consumers' (first causal connection in the strong condition) were compared with the scores for 'Playing Mozart in home furnishing shops means consumers will stay there longer and listen to the music' (first causal connection in the weak condition). Then, the scores for 'If the mood of consumers in home furnishing shops is more positive, the turnover will be raised' (second causal connection in the strong condition) were compared with the scores for 'If consumers in home furnishing shops stay longer and listen to the music, the turnover will be raised' (second causal connection in the weak condition). I used independent-samples t-tests, as I wanted to compare the mean scores for two different conditions.

If the material had been successfully manipulated, there should have been a significant difference in perceived probability scores between the strong and the weak condition. To 'pass' the manipulation check, I required that either the first or the second causal connection in the strong condition VI would be perceived as *more* probable than the one in the weak condition VII, as long as the other causal connection would not be perceived as *less* probable than the one in the weak condition (equally probable was permitted).

Independent-samples t-tests revealed that the majority of the manipulations were successful (see appendix 6C2). There were two exceptions, in which the manipulation failed. For two texts (texts 7 and 9), it was found that although the second causal connection in the strong condition VI was judged to be *more* probable than the one in the weak condition VII, the first causal condition was considered to be *less* probable (I refer to appendix 6B for texts 7 and 9).

For arguments using scheme (C), there was only one causal connection to be judged. After all, scheme (C) was based on a causal

generalization. One-way analyses of variance were performed to compare the perceived probability scores across the three conditions VIII, IX and X. For example, scores for the following generalizations were compared: 'In most cases, home furnishing shops benefit financially if they play classical music' (strong condition), 'In most cases, home furnishing shops benefit financially if they play popular music' (weak condition) and 'In most cases, home furnishing shops benefit promotionally if they play classical music' (weak condition). Post-hoc tests (LSD) were performed to find out which of the conditions were statistically significant different from one another.

Condition VIII was supposed to be strong, condition IX was supposed to be weak because of a less relevant cause, and condition X was supposed to be weak because of a less relevant effect. So, the intended difference between the conditions concerned the relevance of cause or effect. If a significant difference in persuasiveness would have been found between the strong condition on the one hand and each of the weak conditions on the other, this difference should be due to the intended difference in relevance. However, as the conditions might have possibly differed in the *perceived probability* of the causal connections, the differences in persuasiveness could have also been due to this factor. In other words, the strong condition VIII could have been more persuasive than the weak condition IX and the weak condition X because the causal connection in the strong condition VIII was perceived as more probable by the respondents.

The post-hoc test results (see appendix 6C3) show that in two cases only, the causal connection in the strong condition VIII was perceived as more probable than the causal connection in a weak condition ( $p = .003$  for text 2 and  $p = .019$  for text 4; see appendix 6B for these texts). To be more specific, in both cases it concerned the weak condition X, which aimed to violate the relevant effect criterion. So, it can be said that just in these cases differences in perceived probability might offer a plausible alternative explanation for differences in persuasiveness between the strong condition on the one hand and the weak condition on the other hand. For the better part of the material, however, this alternative explanation appears to be invalid.

### **The argument from example**

Regarding the argument from example, I checked if the examples that were supposed to be typical were indeed perceived as more typical of a category than the examples that were supposed to be atypical of a

category. More specifically, I wanted to know whether the typical example A1 presented in condition XI and the typical example A2 presented in condition XIII were indeed regarded as more typical than the atypical example A1 presented in condition XII and the atypical example A2 presented in condition XV (see Table 6.2). Thus, I was interested in comparisons between specific groups. Univariate analyses of variance with planned comparisons (LSD) were used to compare perceived typicality scores between these groups.

If the material had been successfully manipulated, there should have been differences in mean scores between each of the two typical examples on the one hand and each of the atypical examples on the other hand.

The majority of the material was successfully manipulated (see appendix 6C4). There were two exceptions (text number 5 and 8): examples that were supposed to be perceived differently in terms of typicality were in fact perceived as equally typical (I refer to appendix 6B for texts 5 and 8).

In the next section (6.3.2) reporting the results of the target items, I will indicate (in notes) how the results turn out to be when the data analyses exclude scores on unsuccessfully manipulated items.

### **6.3.2 Target items**

#### **The argument from authority**

Table 6.3 shows the persuasiveness of the strong and weak arguments for the argument from authority.



Table 6.3 Means of persuasiveness scores for the argument from authority with condition as independent variable (*SD* between parentheses; *N* = 186; \* = differs significantly from condition I,  $p < .05$ ).

	(I)	(II)	(III)	(IV)	(V)
Scheme (A)	Strong: meets criteria (1)-(4)	Weak: violates expert criterion (1)	Weak: violates relevant expertise criterion (2)	Weak: violates trustworthi- -ness criterion (3)	Weak: violates opinion-to- conclusion relevance criterion (4)
A says that M leads to E Therefore: M leads to E	4.43 (1.28)	4.34 (1.26)	4.36 (1.29)	4.17* (1.32)	4.24 (1.22)

First, I tested whether there was a main effect for condition. From tests of within-subjects effects, no statistically significant main effect for condition appeared:  $F(4, 740) = 1.65$ ,  $p = .159$ . The interaction effect between condition and case,  $F(4, 740) = .89$ ,  $p = .469$ , was not statistically significant either. This means that when we compare persuasiveness scores across each of the five conditions, all conditions are equally persuasive.

Based on the research question, however, I was interested in conducting specific comparisons: between the strong argument and each of the weak arguments. So, I tested whether the strong argument from authority was more persuasive than each of the weak arguments from authority. A planned comparisons analysis (LSD) revealed a statistically significant difference between the strong argument from authority and the weak argument violating the trustworthiness criterion ( $p = .020$ ). The differences between the strong argument from authority and each of the other weak arguments from authority were not significant ( $p \geq .080$ ), although the difference between the strong argument from authority and the weak argument violating the opinion-to-conclusion relevance criterion may be considered as a trend towards significance ( $p = .080$ ).

In conclusion: if the weak conditions were systematically compared to the condition meeting all criteria, the results suggest that the strong argument from authority is more persuasive than the argument from an untrustworthy authority. There is also a trend towards the strong argument from authority being more persuasive than the argument from an authority whose opinion differs from what is concluded. Using an authority who is not credible as an expert source or not an expert in the relevant field does not make any

difference in persuasiveness, compared with using an authority who meets all standards<sup>105</sup>.

### The argument from cause to effect

Table 6.4 shows the persuasiveness of the strong and weak argument for the argument from cause to effect, based on two causal connections (scheme (B)) or based on a causal generalization (scheme (C)).

Table 6.4 Means of persuasiveness scores for the argument from cause to effect with condition as independent variable (SD between parentheses; scheme (B): N = 187; scheme (C): N = 186; \* = differs significantly from condition VI/VIII,  $p < .01$ ).

	(VI)	(VII)	
Scheme (B)	Strong: meets cause sufficiency criterion (5)	Weak: violates cause sufficiency criterion (5)	
M leads to G G leads to E Therefore: M leads to E	4.33 (1.28)	3.97* (1.43)	
	(VIII)	(IX)	(X)
Scheme (C)	Strong: meets criteria (6)&(7)	Weak: violates cause relevance criterion (6)	Weak: violates effect relevance criterion (7)
Measures like M generally lead to effects like E Therefore: M leads to E	4.01 (1.23)	3.42* (1.41)	3.65* (1.39)

First, I tested for scheme (B) whether there was a main effect for condition. Tests of within-subjects effects showed a statistically significant main effect for condition  $F(1, 186) = 9.89$ ,  $p = .002$  (partial eta squared = .050). The interaction effect between condition and case was not statistically significant ( $F < 1$ ).

<sup>105</sup> I conducted the analysis of variance again, excluding the scores on the improperly manipulated items. Although this time a main effect for condition ( $F(4, 428) = 5.04$ ,  $p = .001$ , partial eta squared = .045) and a main effect for case ( $F(1, 107) = 4.78$ ,  $p = .031$ , partial eta squared = .043) appeared from the overall test, a planned comparisons analysis (LSD) revealed similar results. Again, the analysis revealed a statistically significant difference between the strong argument from authority and the weak argument violating the trustworthiness criterion ( $p = .000$ ). The differences between the strong argument from authority and each of the other weak arguments from authority were, again, not significant ( $p \geq .128$ ).

These results indicate that the strong argument from cause to effect using scheme (B) is more persuasive than the weak argument violating the cause sufficiency criterion. The influence of argument quality on persuasiveness did not depend on whether it was the first or the second representation of a condition (case). In other words, if people judge an argument based on two causal connections, they are more persuaded by a sufficient cause than by an insufficient cause<sup>106</sup>.

Second, I tested for scheme (C) whether there was a main effect for condition. There appeared to be one:  $F(2, 370) = 11.20, p = .000$ , (partial eta squared = .057). There was no significant interaction effect between condition and case:  $F(2, 370) = .43, p = .652$ . This indicates that persuasiveness differed across the three conditions VIII, IX and X. The difference in persuasiveness, however, did not depend on whether it was the first or the second representation of a condition (case).

Before conducting the experiment, I planned to make two specific comparisons: between the strong argument and each of the two weak arguments. A planned comparisons analysis revealed a statistically significant difference in persuasiveness between the strong argument and the weak argument violating the cause relevance criterion ( $p = .000$ ) and also between the strong argument and the weak argument violating the effect relevance criterion ( $p = .002$ ).

The results suggest that if people judge an argument based on a causal generalization, they are more persuaded by a generalization with a relevant cause and a relevant effect than by a generalization with a less relevant cause or by a generalization with a less relevant effect.

### **The argument from example**

Table 6.5 shows the persuasiveness of different conditions for the argument from example, using one example (scheme (D)) or using two examples (scheme (E)).

---

<sup>106</sup> I conducted the analysis of variance again, excluding the scores on the improperly manipulated items. Again, tests of within-subjects effects showed a statistically significant main effect for condition  $F(1, 69) = 22.78, p = .000$  (partial eta squared = .248). The interaction effect between condition and case,  $F(1, 69) = 1.42, p = .238$ , was not statistically significant. These results are comparable to the results including scores on the improperly manipulated items, only the effect size is larger (compare partial eta squared values .248 and .050).

Table 6.5 Means of persuasiveness scores for the argument from example with condition as independent variable (*SD* between parentheses; *N* = 187).

	(XI)		(XII)
Scheme (D)	Strong: meets typicality criterion (8) violates number of ex. crit. (9)		Weak: violates typicality criterion (8) violates number of ex. crit. (9)
In A1, M leads to E Therefore: M leads to E	3.60 (1.33)		3.55 (1.42)
	(XIII)	(XIV)	(XV)
Scheme (E)	Strong: meets typicality criterion (8) meets number of ex. crit. (9) violates diversity criterion (10)	Strong: A2 violates typicality crit. (8) meets number of ex. crit. (9) meets diversity criterion (10)	Weak: violates typicality criterion (8) meets number of ex. crit. (9) violates diversity criterion (10)
In A1, M leads to E In A2, M leads to E Therefore: M leads to E	3.96 (1.45)	3.83 (1.33)	3.70 (1.44)

First, I tested whether there was a main effect of the number of examples. There was a statistically significant difference between using one example and using two examples:  $F(1, 186) = 10.27, p = .001$  (partial eta squared = .052; one-tailed). I also wanted to know whether there was a main effect of typicality. The difference between the two levels of typicality (typical - not typical) was also significant:  $F(1, 186) = 3.20, p = .038$  (partial eta squared = .017; one-tailed). The results almost showed a significant main effect for case; the values nearly reached statistical significance ( $F(1, 186) = 3.88, p = .050$ , partial eta squared = .020). No significant interaction effects were identified ( $p \geq .270$ ).

These results indicate that presenting two examples is more persuasive than presenting one example and that typicality is more persuasive than atypicality. It was also suggested that the persuasiveness depended on whether the respondents were confronted with the first or the second representation of a condition: the first representation was more persuasive than the second. However, it should be noted that the  $p$ -value concerned was exactly equal to .050.

Second, I conducted an analysis for the argument from two examples separately (scheme (E)). I tested whether there were any differences in persuasiveness for the three conditions: two typical examples (not diverse), one typical and one atypical example (diverse) and two atypical examples (not diverse). No main effect of condition was found:  $F(2, 372) = 1.92, p = .147$ . A main effect of case or an interaction effect between condition and case were not found either.

I conducted follow-up tests (planned comparisons) to look at the results for each of the sublevels separately. After all, I wanted to know whether the typical – typical (not diverse) variant and the typical – atypical (diverse) variant were superior to the atypical – atypical (not diverse) variant. The argument containing two typical examples (not diverse) did not significantly differ from the argument containing one typical and one atypical example (diverse) ( $p = .359$ ) and was nearly significantly different from the argument with two atypical examples (not diverse) ( $p = .050$ ). It also appeared that the typical – atypical variant (diverse) and the variant with two atypical examples (not diverse) were equally persuasive ( $p = .299$ ).

Again, I conducted an analysis for the argument from two examples separately (scheme (E)), excluding the scores on the improperly manipulated items. I tested whether there were any differences in persuasiveness for the three conditions: two typical examples (not diverse), one typical example and one atypical example (diverse), and two atypical examples (not diverse). This time, a main effect of condition appeared:  $F(2, 76) = 3.21, p = .046$  (partial eta squared = .078). I also conducted follow-up tests (planned comparisons) to look at the results for each of the sublevels separately. Again, the argument from two typical examples (not diverse) did not significantly differ from the argument from one typical and one atypical example (diverse) ( $p = .640$ ) and the difference with the argument containing two atypical examples (not diverse) revealed a trend towards significance ( $p = .077$ ). This time, however, the results showed that the variant with one typical and one atypical example (diverse) was *more* persuasive than the variant with two atypical examples (not diverse) ( $p = .016$ ). So, if we disregard the scores on the inadequately manipulated material, it appears more persuasive to use one typical example and one atypical example, by which diversity is established, than to use two atypical examples, by which the diversity criterion is violated. If the scores on the inadequate material are taken into consideration, this significant difference between conditions disappears.

In sum: if we consider the results for scheme (E) separately and disregard the scores on the inadequately manipulated material, it is suggested that using two typical examples, which violates the diversity criterion, appears more persuasive than using two atypical examples, which also violates the diversity criterion, although it should be kept in mind that only a trend towards significance was revealed. Compared to the use of two atypical examples, which violates the diversity criterion, presenting one typical and one atypical example, by which diversity is established, is more persuasive.

## **6.4 Conclusion and discussion**

### **Conclusion**

This study addressed the research question for the argument from authority, the argument from cause to effect and the argument from example whether or not normatively strong arguments are more persuasive than normatively weak arguments.

First, the argument from authority. The normatively strong argument from authority was more persuasive than the argument from an untrustworthy authority. Compared to the normatively strong argument, using an authority who is less credible as an expert source did not make any difference in persuasiveness. Using an authority who is an expert in the wrong field did not make a difference either. The difference between the strong argument from authority and the argument from an authority whose opinion differs from what is concluded, almost reached the conventional significance level.

Second, the argument from cause to effect: the normatively strong argument was consistently more persuasive than the normatively weak argument. If people evaluated argumentation based on two causal connections, they were more persuaded by a sufficient cause than by an insufficient cause. If they evaluated argumentation based on a causal generalization, they were more convinced by a relevant cause and effect than by a less relevant cause or a less relevant effect.

Last, the argument from example. Overall, two examples were more persuasive than one and typical examples were more persuasive than atypical examples. Furthermore, looking at the data for the argument containing two examples specifically and disregarding the scores on inadequately manipulated material, the difference between two typical examples and two atypical examples (in either case violating the diversity criterion) approached the

conventional level of significance. In addition, using one typical example and one atypical example, by which diversity is established, is more persuasive than an argument containing two atypical examples, by which the diversity criterion is violated.

## Discussion

Normatively strong arguments are not necessarily more persuasive than normatively weak arguments. It depends on the particular argument type that is used and the criterion that is violated in the normatively weak argument. Summarized:

- Violating the trustworthiness criterion had a negative impact on the actual persuasiveness of the argument from authority. Violating the expert criterion or the relevant expertise criterion did not. Violating the opinion-to-conclusion relevance criterion appeared to have a negative impact on the actual persuasiveness; a trend toward significance was revealed.
- Violating the cause sufficiency criterion, the cause relevance criterion or the effect relevance criterion had a negative impact on the actual persuasiveness of the argument from cause to effect.
- Violating the number of examples criterion and the typicality criterion had a negative impact on the actual persuasiveness of the argument from example. Violating the diversity criterion had a negative impact on the actual persuasiveness of the argument from example, only if both homogeneous examples presented were also atypical.

In this study, the arguments from authority that violated the expert criterion and the relevant expertise criterion were just as persuasive as the strong argument from authority. This outcome does not correspond with earlier findings that for Dutch participants, strong expert evidence is more persuasive than weak expert evidence - created by changing the relevant field of expertise into a less relevant field of expertise (Hornikx, 2005a; Hornikx & Hoeken, 2007). The results also contradict the conclusion of Kruglanski, Raviv, Bar-Tal, Raviv, Sharvit, Ellis et al. (2005) that in social judgment, "the hierarchy of epistemic authorities matters" (p. 369). The preliminary study reported in chapter 5 revealed that respondents *were* sensitive to differences in relevant expertise, but only when a normative claim was defended. When a descriptive claim was defended, the results

were similar to those of the current study: violating the trustworthiness criterion made a difference in judgment, whereas violating the relevant expertise criterion did not.

A possible explanation for the difference between the current study's results and those of Hornikx (2005a) is formed by the characteristics of the respondents. In Hornikx (2005a), the respondents were university students. In the current study, the respondents varied more in age and level of education. So, it is possible that university students, who are confronted regularly with the hierarchy of epistemic authorities during their study, are more sensitive to differences in expertise and field relevance. However, it can be argued that this is not a sufficient explanation, as in the preliminary study reported in chapter 5, the respondents were not university students either, but still showed sensitivity to differences in relevance of expertise. However, on that study it can be commented that the manipulations were less subtle than the ones in the current study, because both expertise and field relevance were manipulated at the same time. This combination may have been more obvious than the more subtle manipulations in the current experiment. In the current study, it appears from the manipulations checks that in general, respondents noticed the intended differences in argument quality. Still, it is possible that they only became aware of these manipulations when answering the manipulation check items.

The preliminary experiment showed that the nature of the claim supported by the argument from authority might play a role in the relationship between argument quality and respondents' judgments. More specifically, when a descriptive claim was presented, the untrustworthiness of the source had a negative effect on judgments, whereas the irrelevance of expertise did not influence judgments. When a normative claim was presented, there was no such interaction between criterion and argument quality. In this experiment, a similar pattern was found: it does matter if the source of the descriptive assertion is untrustworthy, but it does not matter if the descriptive expression comes from someone with lower expert credibility or less relevant expertise. For future research on the effect of message source properties on persuasion, I suggest that the nature of the claim (descriptive versus normative) should be studied more carefully. To my knowledge, such effect studies have not yet been attempted (I refer to Pornpitakpan, 2004, for a review of the researched variables interacting with source credibility). It can also be argued that claim type should be considered in empirical research, based on the distinction that is made in theory between different



types of authorities and the kinds of opinions these are associated with (see also chapter 2).

An alternative explanation for the results is the lack of prior beliefs. I intentionally selected claims for the main test that were not extremely acceptable or unacceptable, to prevent prior beliefs effects. This might have caused a certain modesty among the respondents who were confronted with sources low in expert credibility or relevance of expertise. The following thoughts might have entered their minds: 'This source is clearly not a real expert/an expert with relevant expertise, but who am I – somebody without any relevant knowledge concerning this matter - to reject the source's opinion? At least this source has been thinking about the issue'). Relying on an untrustworthy authority or less relevant opinions, however, might have been a bridge too far, even for those without any prior beliefs.

Violating the numbers of examples criterion, the typicality criterion and the diversity criterion in the argument from example appeared to have a negative impact on actual persuasiveness. This is in line with findings in psychological empirical research. According to Heit (2000), reviewing psychological studies on inductive reasoning over 25 years, the main findings are that number of cases and diversity of cases promote induction if people make inferences from multiple cases, although the results are particularly variable for diversity effects (p. 579). If people make inferences from single cases, the typicality of the premise category promotes induction. Sloman and Lagnado (2005) characterize the psychological phenomena of typicality and diversity as follows:

Typicality: "The more typical premise categories are of the conclusion category, the stronger is the argument. For example, people are more willing to project a predicate from robins to birds than from penguins to birds because robins are more typical birds than penguins" (pp. 102-103).

Diversity: "The less similar premises are to each other, the stronger the argument tends to be. People are more willing to draw the conclusion that all mammals love onions from the fact that hippos and hamsters love onions than from the fact that hippos and rhinos do because hippos and rhinos are more similar than hippos and hamsters" (p. 103).

More recently, Rhodes, Brickman and Gelman (2008) investigated how people evaluate whether limited samples of evidence provide a good basis for induction. They found that young children valued item typicality, whereas older children and adults valued overall sample

diversity. So, according to previous psychological studies, there is a significant amount of evidence that people are guided by number, typicality and diversity in inductive reasoning, something that also appears from the results of this study.

In this experiment, I used criteria from both scholars and laymen in the field of argumentation to create differences in argument quality and I investigated the impact of meeting or violating criteria on actual persuasiveness. The results suggest that the normatively strong arguments are not necessarily more persuasive than the normatively weak arguments. It depends on the particular criterion that is violated in the normatively weak argument.



## **7. General conclusion and discussion**

In this chapter, I will first review the results of the studies presented in this dissertation and answer the research questions that were raised in the introduction (chapter 1). I will discuss the results (7.2) and suggest further research (7.3).

### **7.1 Conclusion**

The aim of this dissertation was to provide insight into the specific characteristics that determine the quality of the argument from authority, the argument from cause to effect and the argument from example in support of a probability claim. This was necessary, because the concept of argument quality in the Elaboration Likelihood Model (ELM) had been insufficiently specified.

I focused on the argument from authority, the argument from cause to effect and the argument from example, because these three types of argument can be used to support the pragmatic argument: a type of argument that is often used in ordinary argumentative practice as well as in persuasion research.

A combination of a normative and a descriptive approach to argumentation was considered a valuable framework for studying argument quality. In normative argumentation studies, criteria have been formulated for the assessment of argument quality. These criteria may be of a more general nature, but they may also be attuned to the specific characteristics of certain types of argument. Descriptive, more empirically based argumentation studies investigate to what extent norms that are formulated on a more theoretical-analytical basis are part of laymen's argumentative competence. The role these norms play in the persuasion process is also examined.

Part I of this dissertation investigated to what extent laymen criteria for evaluating the argument from authority, the argument from cause to effect and the argument from example correspond to criteria formulated in argumentation theory for these argument types. The ELM suggests that people who engage in central processing make a critical evaluation of message arguments. That implicates that they use evaluation criteria to distinguish stronger from weaker arguments. However, it is unclear what these criteria are and how these relate to criteria that have been suggested in normative argumentation theory. Therefore, in part I, I addressed the following questions:

- Research question 1: To what extent do laymen criteria for the argument from authority, the argument from cause to effect and the argument from example correspond to those formulated in argumentation theory?
- Research question 1a: What criteria have been formulated in argumentation theory to evaluate the quality of the argument from authority, the argument from cause to effect and the argument from example?
- Research question 1b: What criteria do laymen use to evaluate the quality of the argument from authority, the argument from cause to effect and the argument from example?

Research question 1a was addressed by means of a literature study. From a selection of publications in the field of argumentation, I made an inventory and a classification of criteria formulated for the evaluation of the three types of argument. Research question 1b was addressed by conducting individual interviews and focus groups with laymen. In these settings, laymen were stimulated by cases to reflect upon the criteria they use to assess argument quality.

The answer to research question 1 is: laymen use scheme-specific criteria and these scheme-specific criteria are generally similar to those identified within normative, argumentation-theoretical literature. However, it appears that respondents do not agree on all criteria to the same extent. Some criteria were mentioned by the majority of the respondents (i.e. agreement is relatively high for some criteria), whereas other criteria were only mentioned by a few of them (i.e. agreement is relatively low for other criteria). For the argument from authority, there is relatively high agreement on the relevant expertise criterion, the expert criterion, the trustworthiness criterion, the ability to provide evidence criterion, the recency criterion and the hedged statement criterion. The first five criteria correspond to the criteria identified in argumentation theory, whereas the last mentioned criterion does not. For the argument from cause to effect, there is relatively high agreement on the cause sufficiency criterion, which has also been identified as an important criterion in argumentation theory. For the argument from example, considerable agreement exists on the relevance criterion, the number of examples criterion, the example accuracy criterion and the causality criterion. The first three criteria correspond to criteria formulated in argumentation theory; the last mentioned criterion does not. About the criteria for

which considerable agreement exists, it appears likely that they are part of the laymen's toolkits in assessing argument quality.

Some evaluation criteria that have been suggested in argumentation theory are *not* used by laymen, but this can in most cases be explained by the material that did not give respondents any reason to use these criteria. In case irrelevant criteria were used by respondents, i.e. criteria that were in fact related to a different argument scheme than the one under consideration, this was often understandable on the basis of the fact that the context created in the material was pragmatic in nature.

According to the ELM, normatively strong arguments should be more persuasive than normatively weak arguments in conditions in which messages are processed centrally. However, empirical research indicates this is not necessarily the case (see, e.g., Hoeken, 2001a). Therefore, the following research question was addressed in part II of this dissertation:

Research question 2: Are normatively strong arguments from authority, arguments from cause to effect and arguments from example actually more persuasive than normatively weak arguments?

Before this research question was addressed, a pilot experiment was carried out to investigate the relation between the perceived reasonableness and the actual persuasiveness of normatively strong and weak arguments. In this study, one group of respondents judged to what extent they considered an argument good support for the conclusion (indicating perceived reasonableness), whereas another group of respondents indicated to what extent they agreed with the conclusion (indicating actual persuasiveness). The argument types concerned were arguments from authority and arguments from cause to effect, varying in argument strength as they did or did not meet certain normative evaluation criteria. The pattern of results was similar for perceived reasonableness and actual persuasiveness. Therefore, in the next experiment, addressing research question 2, the decision was made to measure one dependent variable only: the actual persuasiveness.

This experiment, much like the pilot study, measured to what extent respondents agreed with the conclusion (indicating actual persuasiveness). The argument types concerned were the argument from authority, the argument from cause to effect and the argument from example. For each argument type, a specific set of criteria was

selected to manipulate argument strength. Generally, a strong argument was one meeting certain normative evaluation criteria. Apart from this strong argument, weak versions were created that violated (at least) one criterion and met (all) other criteria.

The answer to research question 2 is: normatively strong arguments from cause to effect and arguments from example are indeed more persuasive than the normatively weak variants, but normatively strong arguments from authority are not always more persuasive than the normatively weak variants. More specifically: for the argument from cause to effect, the violation of all selected criteria mattered: violating the cause sufficiency criterion, the cause relevance criterion or the effect relevance criterion had a negative impact on actual persuasiveness. For the argument from example as well, the violation of all selected criteria appeared to make a difference in persuasiveness: violating the typicality criterion, the number of examples criterion or the diversity criterion had a negative impact on actual persuasiveness, although violating the last mentioned criterion only mattered if both homogeneous examples were also atypical. For the argument from authority, the violation of two of four selected criteria appeared to have a negative impact on actual persuasiveness: the trustworthiness criterion and the opinion-to-conclusion relevance criterion (for the former criterion, there was a significant effect, but for the latter criterion only a trend toward significance was revealed). Violating the expert criterion or the relevant expertise criterion did not make any difference in persuasiveness. These results indicate that laymen use scheme-specific criteria in conditions in which they process arguments centrally. They are less persuaded by arguments from cause to effect and arguments from example that violate these criteria. Arguments from authority that violate some of these criteria, however, can still persuade them.

The results of the main experiment generally correspond to the results of the pilot experiment, except that respondents in the pilot study were sensitive to differences in relevant expertise in the argument from authority. This difference, however, can be explained by the fact that in the pilot study, the manipulations of expertise included both the level of education as the relevance of the expertise, whereas in the main experiment, expertise was only manipulated through violating the relevant expertise criterion.

All in all, it may be concluded that the ELM's prediction about normatively strong arguments being more persuasive than normatively weak arguments during central processing, can not be

fully confirmed by the results of this study. Indeed, normatively strong arguments from cause to effect and arguments from example are actually more persuasive than the normatively weak variants. However, about the argument from authority, it can be said that what should be strong according to argumentation theory and to laymen, is not necessarily more persuasive.

## **7.2 Discussion**

In this section, I will present explanations for the findings in this dissertation (7.2.1) and discuss the implications of the results for persuasion research and argumentation studies (7.2.2).

### **7.2.1 Explanations**

Although laymen use certain evaluation criteria for the evaluation of the argument from authority, they may still be persuaded by arguments from authority that violate these evaluation criteria. In this study, there appears to be a discrepancy between the specific norms for the argument from authority that laymen are aware of and what they are actually persuaded by. This observation is not in accordance with the ELM's hypothesis that normatively strong arguments lead to greater persuasion than normatively weak arguments.

The first explanation is that the task that respondents performed to have their evaluation criteria uncovered, differed from the task respondents performed in the experimental study. To uncover laymen criteria, respondents were presented with cases. They were asked to compare arguments, rank arguments and explain why they considered certain arguments stronger than others. This type of task might have caused the respondents to pay much attention to the quality of the argument. In other words: respondents probably elaborated highly upon argument quality. In the experimental study, on the other hand, respondents judged the extent to which they agreed with the conclusion at hand; they were not asked to compare arguments with respect to quality. Although it is likely that much attention was paid to this task as well, it was up to the respondent to focus on the quality of the argument, the acceptability of the claim, or on both. This may have decreased the chance that respondents indeed elaborated upon argument quality. It is possible that in the experimental situation, in which the respondent's attention was not necessarily drawn towards the quality of the argument, certain manipulations became too subtle, even though the evaluation



criteria concerned appeared from the studies in which the case-task was used.

It is also possible that some weaknesses in arguments are harder to identify than other weaknesses, so that more effort is required in spotting the weaknesses that are hard to identify. This might explain why laymen in this study could still be persuaded by arguments from authority violating the expert criterion and the relevant expertise criterion; it may have been more difficult for respondents to see that a source had lower expert credibility or that the expertise was less relevant, than to notice untrustworthiness or a less relevant opinion (the degree of elaboration being equally high).

An alternative explanation is that although laymen are aware of criteria they should use in their evaluations of arguments, they find some criteria more important than other criteria. For example: it is possible that people are aware that sources need to have relevant expertise and should be trustworthy, but that they only apply the trustworthiness norm, as they regard this norm as relatively more important than the other criterion they know. An analogy can be drawn with someone evaluating a letter and only focussing on typing errors and not on other aspects, like text structure, or grammatical errors, although he or she knows that these are also aspects to pay attention to.

Finally, it is possible that it can be difficult for people to test a certain argument against all critical tests that are stored in memory. Evaluating the quality of an argument may be very a demanding task: one needs to read the text, to understand what is written, to form a representation in the mind, to identify the type of argument, to recall his or her own beliefs, to retrieve the relevant evaluation criteria from memory and to use them. In that demanding situation, cognitive resources are perhaps too limited to recall and use all critical questions. In addition, if the particular task is to judge the acceptability of the claim, as in the experimental study, people might be less tempted to use their cognitive resources for all these subtasks and might prefer focussing on claim acceptability.

### **7.2.2 Implications for argumentation studies and persuasion research**

In argumentation studies, research has been done on whether discussion moves that are considered fallacies in (pragma-dialectical) theory, are considered unreasonable by laymen in the field of argumentation (Van Eemeren et al., 2007). However, more insight

was needed into the conventional validity of normative-theoretical criteria pertaining to specific argument schemes. The results of the current study indicate that normative-theoretical criteria and laymen criteria correspond to a considerable extent, although laymen do not report on all criteria to the same extent. Hence, it appears as if, similar to fallacies, the criteria that should be used to assess argument quality are part of laymen's argumentative competence.

However, if we take the actual persuasiveness of normatively strong and weak arguments into consideration, the conventional validity of evaluation criteria appears limited. In a situation in which laymen compare arguments, they mention criteria that also occur in literature. However, in an experimental situation, they are persuaded by arguments that do not meet these criteria. Therefore, the conventional validity of evaluation criteria can, in a generalizing sense, only for a part apply to the setting in which people judge the acceptability of the claim supported by arguments that do or do not meet these criteria.

Second, the concept of central processing in the ELM can now be further specified. The studies reported here indicate that if people carefully examine arguments from authority, arguments from cause to effect and arguments from example, they are able to apply evaluation criteria related to these types of argument. Schellens and De Jong (2004) claimed, on the basis of the patterns of argumentation found in persuasive brochures, that readers who process these texts centrally will have to be able to ask and answer relevant evaluation questions related to certain types of argument, among which the argument from cause to effect, the argument from example and the argument from authority. The results of this study show that people who are engaged in central processing are indeed able to do so.

Third, there has repeatedly been a call for a conceptually meaningful definition of argument quality and better insight into the specific characteristics that determine argument strength (see, e.g., Areni, 2003; O'Keefe, 1995, 2002, 2008; Petty & Cacioppo, 1986a). The studies reported here show that it will be a difficult mission to develop one meaningful definition of argument quality. After all, argument quality depends on what type of argument is used, and by defining the specific characteristics that determine argument strength, one has to take into consideration that what is strong in theory, is not necessarily considered strong by laymen. In addition, the characteristics that laymen claim make up a strong argument do not necessarily correspond to characteristics that persuade them. In testing ELM-predictions concerning the relationship between

argument quality and persuasive effects, it is important that one selects a type of argument for which the evaluation criteria of lay people are known and to manipulate precisely these criteria. Under conditions of central processing, the ELM would predict that people apply these criteria, and on the basis of the outcome of the evaluation process decide upon the strength of the argument. The cognitive responses generated by centrally processing participants should reveal the employment of these criteria.

Fourth, the introduction stated that differences in argument quality also play an important methodological role in persuasion research, as they can be used to assess whether respondents have been scrutinizing the message or not. Only participants who pay close attention to the arguments will notice the difference between messages containing strong arguments and those containing weak arguments. If differences in argument quality are used as a methodological tool, it is important to create differences that are likely to be noticed by participants. Therefore, it is relevant that differences in argument quality are not only based on theoretical considerations, but also on knowledge of what laymen believe is normatively strong and normatively weak. As the results of this study provide that kind of knowledge, they can be regarded as a contribution to methodology in persuasion research.

Scholars with a normative approach to argumentation studies should also reconsider their typologies of argument schemes with respect to the argument from example, for which it appears wise to distinguish between subtypes that ask for different evaluation questions: the example to a descriptive generalization and the example to a causal generalization. In addition, some evaluation criteria may need some refinement, such as the strictly formulated sufficiency criterion for the argument from cause to effect, or the loosely defined criterion referring to typicality/representativeness (pertaining to the argument from example).

### **7.3 Limitations and suggestions for further research**

#### **Part I**

##### **Reliability**

An assessment was made on the extent to which raters agreed on the criteria that should be attached to certain phrases within a fragment (see chapter 2, 3 and 4). However, the question remains to

what extent agreement would have existed on other steps in the data-analysis, such as dividing transcripts into fragments, deciding whether or not a respondent uses a certain evaluation criterion, and the grouping of criteria on the basis of a shared characteristic.

The sample size used in the reliability tests may pose an additional problem. According to Neuendorf (2002), there is no set standard for the number of units that should be used in a reliability assessment, but in general, the reliability sample should probably never be smaller than 50 and should rarely need to be larger than about 300 (p. 159). However, only in the study on the argument from authority (chapter 2), the sample size exceeded 50 ( $n = 79$ ). In the studies on the argument from cause to effect and the argument from example, sample sizes were 12 and 15, respectively.

In determining reliability, I relied on Peat and Barton's (2005) guidelines. According to Neuendorf (2002), however, it is open to debate what should be an acceptable level of intercoder reliability. From a review of the work on reliability, she concludes "that reliability coefficients of .90 or greater would be acceptable to all, .80 or greater would be acceptable in most situations, and below that, there exists great disagreement. In general, the beyond-chance statistics, such as Scott's *pi* and Cohen's *kappa*, are afforded a more liberal criterion" (p. 143). Relying on this conclusion, the value of .804 found in the study on the argument from authority would (in general) be acceptable, but the acceptability of values found in the other studies (.636 for the argument from cause to effect; .634 for the argument from example), would probably be cause for discussion.

## Validity

First, comments can be made on the cases that were used to uncover laymen criteria. The open cases used in the study on the argument from authority (chapter 2) appeared to be more successful in uncovering laymen criteria than the open cases used in the studies on the argument from cause to effect (chapter 3) and the argument from example (chapter 4). In the studies on the argument from cause to effect and the argument from example, a considerable amount of respondents presented argument types and evaluation criteria that were not intended. Possibly, the task of inviting one suitable guest and one unsuitable guest for a television programme is easier to accomplish than the task of finding a strong and a weak argument from cause to effect/argument from example, especially if it concerns

arguments in support of a claim predicting that a certain consequence will occur as a result of a certain measure.

In addition, respondents that did not come up with the desired argument types were not corrected by the interviewer. For instance, if a respondent presented a pragmatic argument instead of an argument from example, the respondent was not told that he or she should have come up with an argument from example instead of a pragmatic argument. Had the interviewer corrected respondents, the open cases might have been more useful. On the other hand, correcting respondents would have increased the risk of interviewer bias: the interviewer is searching for an answer that supports a preconceived idea. In this case, the preconceived idea would be that a certain argument scheme plays a role in the respondent's mind, whereas in fact that idea may be wrong. It is possible, for instance, that it would never come to mind to use an argument from example to support a causal generalization. So if an interviewer were to tell a respondent to use an argument from example to support a causal generalization instead of a pragmatic argument, this could cause a response effect. In future research, interview techniques should be found that enable the researcher to determine whether or not a respondent has knowledge of a certain argument type.

It should also be emphasized that the evaluation criteria used by the laymen in this study are rarely formulated on the same abstract level as the argumentation-theoretical criteria. Rather, respondents expressed themselves in concrete terms, induced by the material they were to judge. This observation is in line with Van Eemeren et al. (2007) who found that lay people generally do not refer to abstract-general discussion rules but rather indicated in concrete terms why they found a certain discussion move reasonable or unreasonable.

One may also raise the following objection to using cases: the cases are not used to *uncover* laymen criteria, but to *create* laymen criteria. The norms might as well be created during the interviews, in which laymen, possibly for the first time in their lives, were explicitly confronted with differences in argument quality. So, how do we know that laymen already used these norms when they stepped into the door to the research setting? My answer is: we do not know for sure. It is possible that some respondents, by doing the tasks they were requested to do, creating a strong and a weak argument in the open case and by ranking arguments in the closed case, developed their norms during the tasks. This suggestion applies especially to those criteria that were mentioned by a relatively small number of

respondents; the suggestion is less plausible for those criteria that were mentioned by a relatively high number of respondents.

## **Part II**

In the experiments in chapter 5 and 6, no control condition was used. A control condition would be a condition in which only claims are presented to respondents, without any supporting arguments. One function of such a control condition would be to check whether adding arguments would be more persuasive than no arguments at all, in case the strong and the weak condition would not have shown any significant difference. Another function would be to see - in case the strong and the weak condition would differ significantly - whether the strong and the weak condition would lead to significantly higher claim acceptance scores compared with the control condition. The reason for not including control conditions in the designs, was that in the within-subjects designs as used in the experiments, respondents would have probably noticed the difference between the items with and the items without any supporting arguments and as a result, inferred the aim of research (or at least the intended difference between the control condition and the other conditions).

Part II argued that it is important to select claims for the experiments that are neither extremely acceptable nor extremely unacceptable, as what respondents already strongly believe or reject might influence the results. However, the problem of selecting neutral claims is that it decreases the chance that respondents become involved with the topic of the message. This might be particularly problematic in cases where an argument from authority is used, as the source might be believed for the (only) reason that the source is more involved with the topic than the receiver. Furthermore, as according to the ELM topic involvement is expected to increase the chance that arguments are examined attentively, the selection of neutral, 'unloaded' claims might have prevented some respondents from identifying weaknesses in arguments that require deep thinking. In addition, it may be difficult for respondents to stay involved with each new topic presented in the questionnaire. It will be a challenge to create material that maximizes the chance of involvement and minimizes the chance of claims being too acceptable or too unacceptable.

In the experimental studies reported in this dissertation, argumentative texts were used that consisted of a few sentences only. Hoeken and Hustinx (2007) showed that ordinary language

users have difficulties differentiating between strong and weak arguments, if the arguments are presented in a longer text. Therefore, it is uncertain whether or not people are more easily persuaded by normatively weak arguments if these arguments are part of a larger text. In addition, the question remains to what extent the short texts as used in the experiments reflect true argumentation, i.e. argumentation occurring in real-life persuasive documents. On the one hand, by using short texts, influences on the measurements can be well-controlled, but on the other hand, generalizability would be improved by using less artificial, experimental material. Experimenting with argument quality in texts that are likely to occur in real life, without losing experimental control, is a future research challenge that should be addressed.

In future ELM research, the relationship of argument quality and persuasive effects can now be explored directly and empirically. O'Keefe (1995) claimed that this would be possible, once an independently-motivated account of argument quality was offered (p. 15). Obviously, the current study does not offer a universal, all-embracing account of argument quality, but for the argument from authority, the argument from cause to effect and the argument from example, it gives insight into what makes for stronger and weaker arguments, from an argumentation-theoretical perspective as well as lay people's perspectives. Therefore, in subsequent ELM research, these insights may be used for creating experimental material and determining what exact argument characteristics contributed to the effects.

In addition, the results of the current study clarify why normatively strong arguments may lead to greater persuasiveness under conditions of central processing: because people use specific criteria during their careful evaluations, on the basis of which they assess argument strength. If, in future ELM research, participants will be instructed to record their thoughts elicited by a strong or weak message that they have evaluated carefully (see, e.g., Petty & Cacioppo, 1986b), these recorded thoughts should reveal the use of (specific) evaluation criteria. This means that in coding these thoughts, special attention should be paid to whether or not participants used these criteria. This kind of information may not only provide evidence that messages were indeed scrutinized systematically, but may also explain why, subsequently, people react favorably, unfavorably, or neutrally to the standpoint advocated. The

nature of argument quality will remain a topic of study, but prospects should be good for empirical work on persuasion in future.





## References

- Areni, C. S. (2003). The effects of structural and grammatical variables on persuasion: An Elaboration Likelihood Model perspective. *Psychology & Marketing, 20*, 349-375.
- Areni, C. S., & Lutz, R. J. (1988). The role of argument quality in the Elaboration Likelihood Model. In M. J. Houston (Ed.), *Advances in Consumer Research: Vol. 15*. (pp. 197-203). Provo, UT: Association for Consumer Research.
- Arthos, J. (2003). Where there are no rules or systems to guide us: Argument from example in a hermeneutic rhetoric. *Quarterly Journal of Speech, 89*, 320-344.
- Baarda, D. B., Goede, M. P. M., de & Teunissen, J. (1995). *Basisboek kwalitatief onderzoek [Introduction to qualitative research]*. Houten: Educatieve Partners Nederland.
- Bergkvist, L., & Rossiter, J. R. (2007). The predictive validity of multiple-item versus single-item measures of the same constructs. *Journal of Marketing Research, 44*, 175-184.
- Bochenski, J. M. (1974). An analysis of authority. In F. J. Adelman (Ed.), *Authority* (pp. 56-85). The Hague: Martinus Nijhoff.
- Brady, H. E., & Seawright, J. (2004). *Framing social inquiry: from models of causation to statistically based causal inference*. Paper prepared for the American Political Science Association Annual Meetings. Chicago, Illinois. [www.asu.edu/clas/polisci/cqrm/APSA\\_2004/BradySeawright.pdf](http://www.asu.edu/clas/polisci/cqrm/APSA_2004/BradySeawright.pdf). Accessed 21 February 2008.
- Collingwood, R. G. (1938/1991). On the so-called idea of causation. *Proceedings of the Aristotelian society 38*. Reprint in A. B. Schoedinger (Ed.), *Introduction to metaphysics: the fundamental questions* (pp. 145-162). Buffalo, N.Y.: Prometheus Books.
- Copi, I. M. (1982). *Introduction to logic (6th ed.)*. New York: Macmillan.
- De George, R. T. (1976). The nature and function of epistemic authority. In R. B. Harris, (Ed.), *Authority: A philosophical analysis* (pp. 76-93). Alabama: The University of Alabama Press.
- De George, R. T. (1985). *The nature and limits of authority*. Kansas: University Press of Kansas.
- Dillard, J. P., Weber, K. M., & Vail, R. G. (2007a). The relationship between the perceived and actual effectiveness of persuasive messages: A meta-analysis with implications for formative campaign research. *Journal of Communication, 57*, 613-631.
- Dillard, J. P., Shen L., & Vail, R. G. (2007b). Does perceived message effectiveness cause persuasion or vice versa? 17

- consistent answers. *Human Communication Research*, 33, 467-488.
- Eemeren, F. H. van (2001). *Crucial concepts in argumentation theory*. Amsterdam: Amsterdam University Press.
- Eemeren, F. H. van, & Garssen, B. (2007). Kwantitatief empirisch onderzoek van argumentatie: Ter inleiding [Quantitative empirical argumentation research: An introduction]. *Tijdschrift voor Taalbeheersing*, 29, 251-274.
- Eemeren, F. H. van, & Grootendorst, R. (1992). *Argumentation, communication, and fallacies: A pragma-dialectical perspective*. Hillsdale, NJ: Lawrence Erlbaum.
- Eemeren, F. H. van, & Grootendorst, R. (2003). A pragma-dialectical procedure for a critical discussion. *Argumentation*, 17, 365-386.
- Eemeren, F. H. van, & Grootendorst, R. (2004). *A systematic theory of argumentation: The pragma-dialectical approach*. Cambridge: Cambridge University Press.
- Eemeren, F. H. van, & Houtlosser, P. (2006). The case of Pragma-Dialectics. In S. Parsons et al. (Eds.), *ArgMAS 2005, Lecture notes in computer science*, 4049 (pp. 1-28). Berlin/Heidelberg: Springer.
- Eemeren, F. H. van, & Houtlosser, P. (2007). The study of argumentation as normative pragmatics. *Pragmatics & Cognition* 15, 161-177.
- Eemeren, F. H. van, Garssen, B., & Meuffels, B. (2007). De conventionele deugdelijkheid van de pragma-dialectische discussieregels [The conventional validity of the pragma-dialectical discussion rules]. *Tijdschrift voor Taalbeheersing*, 29, 251-274.
- Eemeren, F. H. van, Grootendorst, R., & Snoeck Henkemans, A. F. (1996). *Fundamentals of argumentation theory: A handbook of historical backgrounds and contemporary developments*. Mahwah, N.J.: Lawrence Erlbaum Associates.
- Eemeren, F. H. van, & Snoeck Henkemans, A. F. (2006). *Argumentatie: Inleiding in het analyseren, beoordelen en houden van betogen [Argumentation: Introduction into the analysis, judgment and presentation of arguments]*. Groningen/Houten: Wolters-Noordhoff.
- Ehninger, D., & Brockriede, W. (1963). *Decision by debate*. New York: Dodd, Mead.
- Ehninger, D., & Brockriede, W. (1978). *Decision by debate* (2<sup>nd</sup> ed). New York: Harper & Row.

- Enschot-van Dijk, R. van, Hustinx, L., & Hoeken, H. (2003). The concept of argument quality in the Elaboration Likelihood Model: A normative and empirical approach to Petty and Cacioppo's 'strong' and 'weak' arguments. In F. H. van Eemeren, J. A. Blair, C. A. Willard, & A. F. Snoeck Henkemans (Eds.), *Anyone who has a view: Theoretical contributions to the study of argumentation* (pp. 319-335). Dordrecht: Kluwer Academic Publishers.
- Feteris, E. T. (2002). A pragma-dialectical approach of the analysis and evaluation of pragmatic argumentation in a legal context. *Argumentation*, 16, 349-367.
- Fishbein, M., & Ajzen, I. (1981). Acceptance, yielding, and impact: Cognitive processes in persuasion. In R. E. Petty, T. Ostrom, & T. Brock (Eds.), *Cognitive Responses in Persuasion* (pp. 339-359). Hillsdale, NJ: Erlbaum.
- Freeley, A. J. (1993). *Argumentation and debate: Critical thinking for reasoned decision making (8th ed.)*. Belmont, CA: Wadsworth Publishing.
- Freeley, A. J., & Steinberg, L. (2000). *Argumentation and debate: Critical thinking for reasoned decision making (10<sup>th</sup> ed.)*. Belmont, CA: Wadsworth Publishing.
- Garsen, B. J. (1997). *Argumentatieschema's in pragma-dialectisch perspectief: Een theoretisch en empirisch onderzoek [Argumentation schemes in a pragma-dialectic perspective: A theoretical and empirical study]*. Amsterdam: IFOTT.
- Godden, D. M., & Walton, D. (2007). Advances in the theory of argumentation schemes and critical questions. *Informal Logic*, 27, 267-292.
- Godin, B. (1999). Argument from consequences and the urge to polarize. *Argumentation*, 13, 347-365.
- Goodwin, J. (1998). Forms of authority and the real ad verecundiam. *Argumentation*, 12, 267-280.
- Greenbaum, T. L. (1998). *The handbook for focus group research*. Thousand Oaks, CA: Sage.
- Hansen, H. V. (2006). Whately on arguments involving authority. *Informal Logic*, 26, 319-340.
- Hartman, J. (2004). Using focus groups to conduct business communication research. *Journal of Business Communication*, 41, 402-410.
- Hastings, A. C. (1962). *A reformulation of the modes of reasoning in argumentation*. Unpublished doctoral dissertation, Northwestern University, Evanston, IL.

- Heit, E. (2000). Properties of inductive reasoning. *Psychonomic Bulletin & Review*, 7, 569-592.
- Hoeken, H. (1997). Een model voor de rol van argumenten in het overtuigingsproces [A model for the role of arguments in the persuasion process]. In H. van den Bergh, D. Janssen, N. Bertens, & M. Damen (Eds.), *Taalgebruik ontrafeld* (pp. 55-66). Dordrecht: Foris.
- Hoeken, H. (2001a). Anecdotal, statistical, and causal evidence: Their perceived and actual persuasiveness. *Argumentation*, 15, 425-437.
- Hoeken, H. (2001b). Convincing citizens: The role of argument quality. In D. Janssen, & R. Neutelings (Eds.), *Reading and writing public documents* (pp. 147-169). Amsterdam/Philadelphia: Benjamins.
- Hoeken, H., & Hustinx, L. (2003). The relative persuasiveness of anecdotal, statistical, causal and expert evidence. In F. H. van Eemeren, J. A. Blair, C. A. Willard, & A. F. Snoeck Henkemans (Eds.), *Proceedings of the fifth conference of the International Society for the Study of Argumentation* (pp. 497-501). Amsterdam: Sic Sat.
- Hoeken, H., & Hustinx, L. (2006). De relatieve overtuigingskracht van anekdotische en statistische evidentie: Welke rol speelt het argumenttype? [The relative persuasiveness of anecdotal and statistical evidence: What role does argument type play?]. In H. Hoeken, B. Hendriks, & P. J. Schellens (Eds.), *Studies in Taalbeheersing 2* (pp. 98-107). Assen: Van Gorcum.
- Hoeken, H., & Hustinx, L. (2007). The influence of additional information on the persuasiveness of flawed arguments by analogy. In F. H. van Eemeren, J. A. Blair, C. A. Willard, & B. Garsen (Eds.) *Proceedings of the sixth conference of the International Society for the Study of Argumentation* (pp. 625-630). Amsterdam: Sic Sat.
- Hoeken, H., & Hustinx, L. (2009). When is statistical evidence superior to anecdotal evidence in supporting probability claims?: The role of argument type. *Human Communication Research*, 35, [to be published/pages unknown].
- Hornikx, J. (2005a). *Cultural differences in the persuasiveness of evidence types in France and the Netherlands*. [S.l]: F&N Boekservice.
- Hornikx, J. (2005b). A review of experimental research on the relative persuasiveness of anecdotal, statistical, causal and expert evidence. *Studies in Communication Sciences*, 5, 205-216.

- Hornikx, J. (2008). Comparing the actual and expected persuasiveness of evidence types: how good are lay people at selecting persuasive evidence? *Argumentation*, 22, 555-569.
- Hornikx, J., & Hoeken, H. (2007). Cultural differences in the persuasiveness of evidence types and evidence quality. *Communication Monographs*, 74, 443-463.
- Hornikx, J., Starren, M., & Hoeken, H. (2003). Cultural influence on the relative occurrence of evidence types. In F.H. van Eemeren, J. A. Blair, C. A. Willard, & A. F. Snoeck Henkemans (Eds.), *Proceedings of the fifth conference of the International Society for the Study of Argumentation* (pp. 531-536). Amsterdam: Sic Sat.
- Hulswit, M. (2002). *From cause to causation: A Peircean perspective*. Dordrecht: Kluwer Academic Publishers.
- Hume, D. (1739-40/1978). *A treatise of human nature*. Oxford: Oxford University Press.
- Kienpointner, M. (1992). *Alltagslogik: Struktur und Funktion von Argumentationsmustern*. Stuttgart-Bad Cannstatt: Friedrich Frommann Verlag.
- Krueger, R. A., & Casey, M. A. (2000). *Focus groups. A practical guide for applied research* (3<sup>rd</sup> ed.). Thousand Oaks, CA: Sage.
- Kruglanski, A. W., Raviv, A., Bar-Tal, D., Raviv, A., Sharvit, K., Ellis, S., et al. (2005). Says who? Epistemic authority effects in social judgment. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology: Vol. 37*. (pp. 346-392). New York: Academic Press.
- Lewis, D. (1973/1993). Causation. *Journal of Philosophy*, 70, 556-567. Reprint in E. Sosa and M. Tooley (Eds.), *Causation* (pp. 193-204). Oxford: Oxford University Press.
- Littejohn, S. W. (2002). *Theories of human communication* (7<sup>th</sup> ed.). Belmont, CA: Wadsworth.
- Mackie, J. L. (1974). *The cement of the universe: A study of causation*. Oxford: The Clarendon Press.
- McBurney, J. H., & Mills, G. E. (1964). *Argumentation and debate. Techniques of a free society*. New York, USA: Macmillan.
- McGuire, W. J. (2000). Standing on the shoulders of ancients: Consumer research, persuasion, and figurative language. *Journal of Consumer Research*, 27, 109-114.
- Meany, J., & Shuster, K. (2002). *Art, argument and advocacy: Mastering parliamentary debate*. New York: The International Debate Education Association.
- Mill, J. S. (1874). *A system of logic* (8<sup>th</sup> ed.). New York: Harper and Brothers.

- Neuendorf, K. A. (2002). *The content analysis guidebook*. Thousand Oaks, CA: Sage.
- O'Keefe, D. J. (1993). Understanding social influence: Relations between lay and technical perspectives. *Communication Studies*, 44, 228-238.
- O'Keefe, D. J. (1995). Argumentation studies and dual-process models of persuasion. In F. H. van Eemeren, R. Grootendorst, J. A. Blair, & C. A. Willard (Eds.), *Perspectives and Approaches: Proceedings of the third ISSA conference on argumentation* (pp. 3-17). Amsterdam: Sic Sat.
- O'Keefe, D. J. (2002). *Persuasion: Theory and research* (2<sup>nd</sup> ed.). Thousand Oaks, CA: Sage.
- O'Keefe, D. J. (2003). Message properties, mediating states, and manipulation checks: Claims, evidence, and data analysis in experimental persuasive message effect research. *Communication Theory*, 13, 251-274.
- O'Keefe, D. J. (2008). Elaboration likelihood model. In W. Donsbach (Ed.), *International encyclopedia of communication, Vol. 4*. (pp. 1475-1480). Oxford, UK, and Malden, MA: Wiley-Blackwell.
- O'Keefe, D. J. (in press). Theories of persuasion. In R. Nabi, & M. B. Oliver (Eds.), *Handbook of media effects*. Thousand Oaks, CA: Sage.
- O'Keefe, D. J., & Jackson, S. (1995). Argument quality and persuasive effects: A review of current approaches. In S. Jackson (Ed.), *Argumentation and Values: Proceedings of the ninth Alta conference on argumentation* (pp. 88-92). Annandale, VA: Speech Communication Association.
- Peat, J. K., & Barton, B. (2005). *Medical statistics: A guide to data analysis and critical appraisal*. Malden, Mass.: Blackwell Publishing Ltd.
- Perelman, C., & Olbrechts-Tyteca, L. (1969). *The New Rhetoric: A treatise on argumentation*. Notre Dame, Indiana: University of Notre Dame Press.
- Perloff, R. M. (2003). *The dynamics of persuasion: Communication and attitudes in the 21th century* (2<sup>nd</sup> ed.). Mahwah, NJ: Lawrence Erlbaum.
- Petty, R. E. (1994). Two routes to persuasion: state of the art. In G. d'Ydewalle, P. Eelen, P. Bertelson (Eds.), *International perspectives on psychological science, Vol. 2*. (pp. 229-247). Hillsdale, NJ: Erlbaum.
- Petty, R. E., & Cacioppo, J. T. (1983). Central and peripheral routes to persuasion: Application to advertising. In L. Percy, & A.

- Woodside (Eds.), *Advertising and Consumer Psychology* (pp. 3-23). Lexington, MA: D. C. Heath.
- Petty, R. E., & Cacioppo, J. T. (1984). The effects of involvement on responses to argument quantity and quality: Central en peripheral routes to persuasion. *Journal of Personality and Social Psychology, 46*, 69-81.
- Petty, R. E., & Cacioppo, J. T. (1986a). *Communication and persuasion: Central and peripheral routes to attitude change*. New York: Springer.
- Petty, R. E., & Cacioppo, J. T. (1986b). The elaboration likelihood model of persuasion. In L. Berkowitz (Ed.), *Advances in experimental social psychology, Vol. 19*. (pp. 123-205). New York: Academic Press.
- Petty, R. E., & Wegener, D. T. (1998). Attitude change: Multiple roles for persuasion variables. In D. Gilbert, S. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology (4<sup>th</sup> ed.)*. (pp. 323-390). New York: McGraw-Hill.
- Petty, R. E., Fabrigar, L. R., & Wegener, D. T. (2003). Emotional factors in attitudes and persuasion. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of Affective Sciences* (pp. 752-772). Oxford: Oxford University Press.
- Petty, R. E., Wegener, D. T., & Fabrigar, L. R. (1997). Attitude and attitude change. *Annual Review of Psychology, 48*, 609-647.
- Pornpitakpan, C. (2004). The persuasiveness of source credibility: A critical review of five decades' evidence. *Journal of Applied Social Psychology, 34*, 243-281.
- Reinard, J. (1991). *Foundations of argument: Effective communication for critical thinking*. Dubuque, IA: Wm. C. Brown.
- Reynolds, R. A., & Reynolds, J. L. (2002). Evidence. In J. P. Dillard, & M. Pfau (Eds.), *The persuasion handbook: Developments in theory and practice* (pp. 427-444). Thousand Oaks, CA: Sage.
- Rhodes, M., Brickman, D., & Gelman, S. A. (2008). Sample diversity and premise typicality in inductive reasoning: Evidence for developmental change. *Cognition, 108*, 543-556.
- Ritchie, J., Lewis, J., & Elam, G. (2003). Designing and selecting samples. In J. Ritchie, & J. Lewis (Eds.), *Qualitative research practice: A guide for social science students and researchers* (pp. 77-108). Thousand Oaks, CA: Sage.
- Romero, A. A. (1999). The logical form of the unimodel: An argument against information equivalence. *Psychological Inquiry: An international journal of peer commentary and review, 10*, 163-166.



- Rossiter, J. R. (2002). The C-OAR-SE procedure for scale development in marketing. *International Journal of Research in Marketing*, 19, 305-335.
- Schellens, P. J. (1985). *Redelijke argumenten: Een onderzoek naar normen voor kritische lezers [Reasonable arguments: A study on criteria for critical reading]*. Dordrecht: Foris.
- Schellens, P. J. (1986). Types of argument and the critical reader. In F. H. van Eemeren, R. Grootendorst, J. A. Blair, & C. A. Willard (Eds.), *Argumentation: Analysis and practices, Proceedings of the conference on argumentation* (pp. 34-44). Dordrecht: Foris.
- Schellens, P. J., & Jong, M. de (2004). Argumentation schemes in persuasive brochures. *Argumentation*, 18, 295-323.
- Schellens, P. J., & Lagerwerf, L. (2003). Principes in pragmatische argumentatie: Een analyse van de discussie over biotechnologie en voedsel [Principles in pragmatic argumentation: An analysis of the discussion on food biotechnology]. In L. van Waes, P. Cuvelier, G. Jacobs, & I. de Ridder (Eds.), *Studies in Taalbeheersing 1* (pp. 416-427). Assen: Van Gorcum.
- Schellens, P. J., & Verhoeven, G. (1994). *Argument en tegenargument: Een inleiding in de analyse en beoordeling van betogende teksten [Argument and counterargument: An introduction to the analysis and evaluation of argumentative texts]*. Groningen: Martinus Nijhoff.
- Shadish, W., Cook, T., & Campbell, D. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.
- Sherif, M., & Hovland, C. I. (1961). *Social judgment: Assimilation and contrast effects in communication and attitude change*. New Haven: Yale University Press.
- Sherif, C. W., Sherif, M. S., & Nebergall, R. E. (1965). *Attitude and attitude change: The social judgment-involvement approach*. Philadelphia: W.B. Saunders Company.
- Sloman, S. A., & Lagnado, D. A. (2005). The problem of induction. In K. J. Holyoak, & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning*. (pp. 95-116). New York: Cambridge University Press.
- Šorm, E., Timmers, R., & Schellens, P. J. (2007). Determining laymen criteria: evaluating methods. In F. H. van Eemeren, J. A. Blair, C. A. Willard & B. Garssen (Eds.), *Proceedings of the sixth conference of the International Society for the Study of Argumentation* (pp. 1321-1329). Amsterdam: Sic Sat.

- Suppes, P. (1970). *A probabilistic theory of causality*. Amsterdam: North Holland Publishing Company.
- Thompson, V. A., Striener, C. L., Reikoff, R., Gunter, R. W., & Campbell, J. I. D. (2003). Syllogistic reasoning time: Disconfirmation disconfirmed. *Psychonomic Bulletin & Review* 10, 184-189.
- Timmers, R. (to appear). Dissertation Radboud Universiteit Nijmegen.
- Timmers, R., Šorm, E., Schellens, P. J., & Hoeken, H. (2008). De redelijkheid en overtuigingskracht van normatief sterke en normatief zwakke argumenten [The reasonableness and persuasiveness of normatively strong and normatively weak arguments]. *Tijdschrift voor Taalbeheersing*, 30, 117-132.
- Tindale, C. W. (2007). *Fallacies and argument appraisal*. Cambridge: Cambridge University Press.
- Verhoeven, G., & Schellens, P. J. (2008). Argumentatieanalyse [Argumentation analysis]. In P. J. Schellens, & M. Steehouder (Eds.), *Tekstanalyse: Methoden en toepassingen* (pp. 119-161). Assen: Van Gorcum.
- Verschueren, N., Schroyens, W., Schaeken, W. & d'Ydewalle, G. (2004). The interpretation of the concepts 'necessity' and 'sufficiency' in forward unicausal relations. *Current Psychology Letters*, 14. Electronic research journal, document retrieved from <http://cpl.revues.org/document433.html>, 27<sup>th</sup> November 2007.
- Walton, D. N. (1989). Reasoned use of expertise in argumentation. *Argumentation*, 3, 59-73.
- Walton, D. N. (1996). *Argumentation schemes for presumptive reasoning*. Mahwah, NJ: Lawrence Erlbaum.
- Walton, D. N. (1997). *Appeal to expert opinion: Arguments from authority*. Pennsylvania: The Pennsylvania State University Press.
- Walton, D. N. (1999). Rethinking the fallacy of hasty generalization. *Argumentation*, 13, 161-182.
- Walton, D. N. (2004). Classification of fallacies of relevance. *Informal Logic*, 24, 71-103.
- Warnick, B., & Inch, E. S. (1989). *Critical thinking and communication: The use of reason in argument*. New York: Macmillan.
- Wegener, D. T. & Petty, R. E. (1996). Effects of mood on persuasion processes: Enhancing, reducing, and biasing scrutiny of attitude-relevant information. In L. L. Martin, & A. Tesser (Eds.), *Striving and Feeling: Interactions among goals, affect, and self-regulation* (pp. 329-362). Mahwah, NJ: Lawrence Erlbaum Associates

- White, P. A. (1990). Ideas about causation in philosophy and psychology. *Psychological Bulletin*, 108, 3-18.
- Willer, S., Ruchatz, J., & Pethes, N. (2007). Zur Systematik des Beispiels. In J. Ruchatz, S. Willer, & N. Pethes (Eds.), *Das Beispiel: Epistemologie des Exemplarischen* (pp. 7-59). Germany: Kulturverlag Kadmos Berlin.
- Wilson, P. (1983). *Second-hand knowledge: An inquiry into cognitive authority*. Westport, Conn.: Greenwood Press.
- Windes, R. R., & Hastings, A. (1969). *Argumentation and advocacy*. New York: Random House.
- Woods, J., & Walton, D. (1974). Argumentum ad Verecundiam. *Philosophy and Rhetoric*, 7, 135-153.
- Wright, G. H. von (1971). *Explanation and understanding*. London: Routledge and Kegan Paul.

**Appendix 2A Material used for uncovering laymen criteria for the argument from authority: open case with desirability claim**

Wij willen je vragen om je in de volgende situatie te verplaatsen: je hebt gesolliciteerd naar een positie in de redactie van het actualiteitenprogramma *Nova*. Je bent nu op sollicitatiegesprek. Ik ben de P&O manager van *Nova*. Je krijgt de volgende casus:

*Nova* wil een special uitzenden met als standpunt:

**Televisiereclame gericht op kinderen is slecht.**

Jouw taak in de redactie is het uitnodigen van de gasten. Je dient twee gasten uit te nodigen, die het beiden met dit standpunt eens zijn:

Eén gast waarvan je denkt dat die dit standpunt goed kan verdedigen,  
en  
Eén gast waarvan je denkt dat die dit standpunt juist niet goed kan verdedigen.

Het doel van deze opdracht bij jouw sollicitatie is erachter komen of je in staat bent onderscheid te maken tussen geschikte en ongeschikte gasten.

Omdat het hier om een denkbeeldige special gaat, kan in principe iedereen (ook mensen die de Nederlandse taal niet machtig zijn of reeds overleden zijn) worden uitgenodigd. Mocht je eventueel geen naam weten van de persoon die je zou willen uitnodigen, maar wel diens functie (bijv. postbode of stucadoor) dan mag je ook een omschrijving geven van die gast.

**Appendix 2B Material used for uncovering laymen criteria for the argument from authority: open case with probability claim**

Wij willen je vragen om je in de volgende situatie te verplaatsen: je hebt gesolliciteerd naar een positie in de redactie van het actualiteitenprogramma *Nova*. Je bent nu op sollicitatiegesprek. Ik ben de P&O manager van *Nova*. Je krijgt de volgende casus:

*Nova* wil een special uitzenden met als standpunt:

**Televisiereclame gericht op kinderen heeft schadelijke effecten.**

Jouw taak in de redactie is het uitnodigen van de gasten. Je dient twee gasten uit te nodigen, die het beiden met dit standpunt eens zijn.

Eén gast waarvan je denkt dat die dit standpunt goed kan verdedigen,  
en  
Eén gast waarvan je denkt dat die dit standpunt juist niet goed kan verdedigen.

Het doel van deze opdracht bij jouw sollicitatie is erachter komen of je in staat bent onderscheid te maken tussen geschikte en ongeschikte gasten.

Omdat het hier om een denkbeeldige special gaat, kan in principe iedereen (ook mensen die de Nederlandse taal niet machtig zijn of reeds overleden zijn) worden uitgenodigd. Mocht je eventueel geen naam weten van de persoon die je zou willen uitnodigen, maar wel diens functie (bijv. postbode of stucadoor) dan mag je ook een omschrijving geven van die gast.

**Appendix 2C Material used for uncovering laymen criteria for the argument from authority: closed case with desirability claim**

Er zijn verschillende manieren om energie op te wekken. Al deze methoden hebben hun eigen voor- en nadelen. Eén van de meest bediscussieerde methoden is het gebruik van kernenergie. Stel dat je het volgende standpunt moet verdedigen:

**Kernenergie is de beste manier om energie op te wekken.**

Om dit standpunt te onderbouwen, kun je een keuze maken uit onderstaande argumenten. Welke van deze argumenten zou je het liefst gebruiken om dit standpunt te verdedigen? Welke argumenten vind je juist erg zwak ter ondersteuning van het standpunt? Zou je een ordening aan kunnen brengen, waarbij 1 het beste argument is en 7 het slechtste argument?

- A. Sommige natuurkundigen van de Vrije Universiteit van Amsterdam vinden kernenergie de beste manier om energie op te wekken.
- B. Robert Oppenheimer, de vader van de atoombom, noemde kernenergie in de jaren 50 al de beste manier om energie op te wekken.
- C. De bekende voetballers Ruud van Nistelrooij, Edgar Davids en Edwin van der Sar vinden dat kernenergie de beste manier van energie-opwekking is.
- D. Ruud van Wijk, de voorzitter van Greenpeace Nederland, is van mening dat er geen betere manier van energie-opwekking bestaat.
- E. Dr. Frank Verbeet, als kernfysicus verbonden aan kerncentrale Borssele, is al jaren van mening dat kernenergie de beste wijze van energie-opwekking is.
- F. Dat kernenergie de beste methode van energie-opwekking is, vindt ook Wil Derben, 17-jarige VMBO-scholier.
- G. Prof. dr. G. Verhoeven, hoogleraar kernfysica aan de Technische Universiteit Delft, vindt dat onder bepaalde omstandigheden met behulp van kernenergie relatief goed energie opgewekt kan worden.

Keuze:

1. \_\_\_\_\_ 2. \_\_\_\_\_ 3. \_\_\_\_\_ 4. \_\_\_\_\_ 5. \_\_\_\_\_ 6. \_\_\_\_\_ 7. \_\_\_\_\_

**Appendix 2D Material used for uncovering laymen criteria for the argument from authority: closed case with probability claim**

Al jarenlang is het broeikas effect een belangrijk item. Er zijn verschillende manieren bedacht om het broeikas effect terug te dringen. Eén van de meest bediscussieerde methoden is het gebruik van kernenergie. Stel dat je het volgende standpunt moet verdedigen:

**Het veel meer gebruikmaken van kernenergie leidt tot een sterke terugdringing van het broeikas effect.**

Om deze stelling te onderbouwen, kun je een keuze maken uit onderstaande argumenten. Welke van deze argumenten zou je het liefst gebruiken om de stelling te verdedigen? Welke argumenten vind je juist erg zwak ter ondersteuning van de stelling? Zou je een ordening aan kunnen brengen, waarbij 1 het beste argument is en 7 het slechtste argument?

- A. Sommige natuurkundigen van de Vrije Universiteit van Amsterdam denken dat met behulp van kernenergie het broeikas effect sterk teruggedrongen kan worden.
- B. Robert Oppenheimer, de vader van de atoombom, noemde kernenergie in de jaren 50 al veelbelovend als een manier om het broeikas effect tegen te gaan.
- C. De bekende voetballers Ruud van Nistelrooij, Edgar Davids en Edwin van der Sar vinden dat met behulp van kernenergie het broeikas effect het beste verminderd kan worden.
- D. Ruud van Wijk, de voorzitter van Greenpeace Nederland, is van mening dat het broeikas effect het best bestreden kan worden door over te stappen op kernenergie.
- E. Dr. Frank Verbeet, als kernfysicus verbonden aan kerncentrale Borssele, is al jaren van mening dat kernenergie de beste methode om het broeikas effect te bestrijden.
- F. Dat kernenergie de beste oplossing is voor het broeikas effect, vindt ook Wil Derben, 17-jarige VMBO-scholier.
- G. Prof. dr. G. Verhoeven, hoogleraar kernfysica aan de Technische Universiteit Delft, vindt dat onder bepaalde omstandigheden kernenergie een zekere bijdrage kan leveren aan het tegengaan van het broeikas effect.

Keuze:

1. \_\_\_\_\_ 2. \_\_\_\_\_ 3. \_\_\_\_\_ 4. \_\_\_\_\_ 5. \_\_\_\_\_ 6. \_\_\_\_\_ 7. \_\_\_\_\_

**Appendix 3A Material used for uncovering laymen criteria for the argument from cause to effect: open case with claim 1**

Ik zou u willen vragen om u zich eens in de volgende situatie te verplaatsen. U hebt gesolliciteerd naar een baan als speechschrijver voor het ministerie van Volksgezondheid. U bent nu op sollicitatiegesprek. Ik ben de manager van de afdeling Personeelszaken en ik moet testen of u goede toespraken kunt schrijven. U krijgt de volgende opdracht:

---

U moet een toespraak schrijven voor de minister. In deze toespraak verdedigt hij de volgende stelling:

**Als we het maximaal toegestane vetgehalte in gehakt van 35 naar 25 procent verlagen, dan zullen de kosten voor de gezondheidszorg dalen.**

Als speechschrijver weet u dat u een standpunt moet ondersteunen met sterke argumenten. Uw taak is nu om twee argumenten te geven die het standpunt ondersteunen. U kiest ervoor om het standpunt te ondersteunen met voorspellingen. Geef

één sterk argument dat het standpunt goed ondersteunt  
 en  
 één zwak argument dat het standpunt juist niet goed ondersteunt.

Nadat u een sterk en een zwak argument heeft gegeven, krijgt u de vraag waarom het ene argument sterker is ter ondersteuning van het standpunt dan het andere argument.

Het doel van deze opdracht bij uw sollicitatie is erachter te komen of u in staat bent onderscheid te maken tussen sterke en zwakke argumenten.

<b>Als we het maximaal toegestane vetgehalte in gehakt van 35 naar 25 procent verlagen, dan zullen de kosten voor de gezondheidszorg dalen.</b>	
A. Het sterke argument:	B. Het zwakke argument:
<b>Want als we het vetgehalte in gehakt verlagen, dan</b> ..... ..... .....	<b>Want als we het vetgehalte in gehakt verlagen, dan</b> ..... ..... .....
De vraag die u straks wordt gesteld, luidt: waarom is argument A sterker ter ondersteuning van het standpunt dan argument B?	



**Appendix 3B Material used for uncovering laymen criteria for the argument from cause to effect: open case with claim 2**

Ik zou u willen vragen om zich eens in de volgende situatie te verplaatsen. U hebt gesolliciteerd naar een baan als speechschrijver voor het ministerie van Volkshuisvesting, Ruimtelijke Ordening en Milieubeheer. U bent nu op sollicitatiegesprek. Ik ben de manager van de afdeling Personeelszaken en ik moet testen of u goede toespraken kunt schrijven. U krijgt de volgende opdracht:

U moet een toespraak schrijven voor de minister. In deze toespraak verdedigt hij de volgende stelling:

**Als er meer groen in en rond nieuwbouwwijken komt, dan zullen de bewoners van die wijken minder gebruik maken van de auto.**

Als speechschrijver weet u dat u een standpunt moet ondersteunen met sterke argumenten. Uw taak is nu om twee argumenten te geven die het standpunt ondersteunen. U kiest ervoor om het standpunt te ondersteunen met voorspellingen. Geef

één sterk argument dat het standpunt goed ondersteunt  
 en  
 één zwak argument dat het standpunt juist niet goed ondersteunt.

Nadat u een sterk en een zwak argument heeft gegeven, krijgt u de vraag waarom het ene argument sterker is ter ondersteuning van het standpunt dan het andere argument.

Het doel van deze opdracht bij uw sollicitatie is erachter te komen of u in staat bent onderscheid te maken tussen sterke en zwakke argumenten.

<b>Als er meer groen in en rond nieuwbouwwijken komt, dan zullen de bewoners van die wijken minder gebruik maken van de auto.</b>	
A. Het sterke argument:	B. Het zwakke argument:
<p><b>Want als de nieuwbouwwijken groener worden,</b></p> <p><b>dan</b>.....</p> <p>.....</p> <p>.....</p>	<p><b>Want als de nieuwbouwwijken groener worden,</b></p> <p><b>dan</b>.....</p> <p>.....</p> <p>.....</p>
<p>De vraag die u straks wordt gesteld, luidt: waarom is argument A sterker ter ondersteuning van het standpunt dan argument B?</p>	

**Appendix 3C Material used for uncovering laymen criteria for the argument from cause to effect: closed case with claim 3**

Telewerken is thuis werken met behulp van een computeraansluiting met het bedrijf. De werknemer doet zijn werk dan thuis achter de computer in plaats van op kantoor. Dit kan een manier zijn om langdurig zieke werknemers weer aan het werk te krijgen, of om te voorkomen dat werknemers langdurig uitvallen. Stel dat u het volgende standpunt moet verdedigen:

**Als werkgevers langdurig zieke medewerkers in de gelegenheid stellen om te telewerken, dan zullen ze eerder terugkeren naar hun gewone werkplek.**

Om dit standpunt te onderbouwen, kunt u een keuze maken uit de onderstaande argumenten. Welke van deze argumenten vindt u sterk ter ondersteuning van dit standpunt? En welke argumenten vindt u juist erg zwak ter ondersteuning van dit standpunt? Zou u de argumenten willen ordenen, waarbij 1 het beste argument is en 5 het slechtste argument?

- A) Langdurig zieke werknemers die vrijaf krijgen zullen over het algemeen sneller terugkeren naar hun gewone werkplek. Dus: als werkgevers langdurig zieke medewerkers in de gelegenheid stellen om te telewerken, dan zullen ze eerder terugkeren naar hun gewone werkplek.
- B) Als langdurig zieke werknemers kunnen telewerken, dan zijn ze minder snel vermoeid. Uiteindelijk keren ze dan sneller terug naar de gewone werkplek. Dus: als werkgevers langdurig zieke medewerkers in de gelegenheid stellen om te telewerken, dan zullen ze eerder terugkeren naar hun gewone werkplek.
- C) Maatregelen om langdurig zieke werknemers thuis te laten werken leiden over het algemeen tot een oplossing van maatschappelijke en individuele problemen. Dus: als werkgevers langdurig zieke medewerkers in de gelegenheid stellen om te telewerken, dan zullen ze eerder terugkeren naar hun gewone werkplek.
- D) Als zieke werknemers hun werk thuis achter de computer kunnen doen, dan hoeven ze niet naar hun werk te reizen. En als ze niet naar hun werk hoeven reizen, dan worden ze minder belast. En als ze minder worden belast door het reizen, dan kunnen ze snel weer een beetje werk verzetten. Dus: als werkgevers langdurig zieke medewerkers in de gelegenheid stellen om te telewerken, dan zullen ze eerder terugkeren naar hun gewone werkplek.
- E) Als zieke werknemers kunnen telewerken, dan zullen ze meer werk verzetten. Bovendien kunnen ze betrokken blijven bij werk en collega's. Uiteindelijk zullen ze dan sneller terugkeren naar hun gewone werkplek. Dus: als werkgevers langdurig zieke medewerkers in de gelegenheid stellen om te telewerken, dan zullen ze eerder terugkeren naar hun gewone werkplek.

Keuze:

1. \_\_\_\_\_ 2. \_\_\_\_\_ 3. \_\_\_\_\_ 4. \_\_\_\_\_ 5. \_\_\_\_\_

**Appendix 3D Material used for uncovering laymen criteria for the argument from cause to effect: closed case with claim 4**

Het streven van de overheid is om zoveel mogelijk mensen bij cultuur te betrekken. Ook kinderen en jongeren worden gestimuleerd om deel te nemen aan het culturele leven, bijvoorbeeld door naar jeugdtheaterscholen te gaan. Bij de jeugdtheaterscholen volgen kinderen en jongeren lessen in theater, beweging en stemvorming. Stel dat u het volgende standpunt moet verdedigen:

**Als er meer geld beschikbaar wordt gesteld voor jeugdtheaterscholen, dan zal de leefbaarheid van de samenleving bevorderd worden.**

Om dit standpunt te onderbouwen, kunt u een keuze maken uit de onderstaande argumenten. Welke van deze argumenten vindt u sterk ter ondersteuning van dit standpunt? En welke argumenten vindt u juist erg zwak ter ondersteuning van dit standpunt? Zou u de argumenten willen ordenen, waarbij 1 het beste argument is en 5 het slechtste argument?

- A) Initiatieven die mensen bij de samenleving betrekken leiden over het algemeen tot een leefbaardere samenleving. Dus: als er meer geld beschikbaar wordt gesteld voor jeugdtheaterscholen, dan zal de leefbaarheid van de samenleving bevorderd worden.
- B) Als jeugdtheaterscholen een financiële impuls krijgen, dan wordt de jeugd beter opgeleid in amateurtheater. Uiteindelijk wordt de samenleving dan leefbaarder. Dus: als er meer geld beschikbaar wordt gesteld voor jeugdtheaterscholen, dan zal de leefbaarheid van de samenleving bevorderd worden.
- C) Het investeren in jeugdtheaterscholen leidt over het algemeen tot een duurzame toekomst. Dus: als er meer geld beschikbaar wordt gesteld voor jeugdtheaterscholen, dan zal de leefbaarheid van de samenleving bevorderd worden.
- D) Als jeugdtheaterscholen financiële ondersteuning krijgen, dan kunnen ze zich artistiek en organisatorisch versterken. En als jeugdtheaterscholen kwalitatief sterker zijn, dan worden kinderen en jongeren daar beter opgeleid. En als ze goed worden opgeleid op het gebied van kunst en cultuur, dan zullen ze zich in het algemeen ook beter ontplooiën. Dus: als er meer geld beschikbaar wordt gesteld voor jeugdtheaterscholen, dan zal de leefbaarheid van de samenleving bevorderd worden.
- E) Als er meer geld beschikbaar wordt gesteld voor jeugdtheaterscholen, dan krijgen meer jongeren de kans om betrokken te worden bij de maatschappij. Bovendien zullen jongeren zich beter ontwikkelen. Uiteindelijk wordt de samenleving dan leefbaarder. Dus: als er meer geld beschikbaar wordt gesteld voor jeugdtheaterscholen, dan zal de leefbaarheid van de samenleving bevorderd worden.

Keuze:

1. \_\_\_\_\_ 2. \_\_\_\_\_ 3. \_\_\_\_\_ 4. \_\_\_\_\_ 5. \_\_\_\_\_

**Appendix 4A Material used for uncovering laymen criteria for the argument from example: open case with claim 1**

Ik zou u willen vragen om u zich eens in de volgende situatie te verplaatsen. U hebt gesolliciteerd naar een baan als speechschrijver voor het ministerie van Onderwijs, Cultuur en Wetenschappen. U bent nu op sollicitatiegesprek. Ik ben de manager van de afdeling Personeelszaken en ik moet testen of u goede toespraken kunt schrijven. U krijgt de volgende opdracht:

---

U moet een toespraak schrijven voor de minister. In deze toespraak verdedigt hij de volgende stelling:

**Als kinderen van jongs af aan worden voorgelezen, dan zullen ze de Nederlandse taal beter beheersen.**

Als speechschrijver weet u dat u een standpunt moet ondersteunen met sterke argumenten. Uw taak is nu om twee argumenten te geven die het standpunt ondersteunen. U kiest ervoor om het standpunt te ondersteunen met voorbeelden. Geef

- één sterk voorbeeld dat het standpunt goed ondersteunt
- en
- één zwak voorbeeld dat het standpunt juist niet goed ondersteunt.

Nadat u een sterk en een zwak argument heeft gegeven, krijgt u de vraag waarom het ene argument sterker is ter ondersteuning van het standpunt dan het andere argument.

Het doel van deze opdracht bij uw sollicitatie is erachter te komen of u in staat bent onderscheid te maken tussen sterke en zwakke argumenten.

<b>Als kinderen van jongs af aan worden voorgelezen, dan zullen ze de Nederlandse taal beter beheersen.</b>	
A. Het sterke argument:	B. Het zwakke argument:
<p><b>Want ik ken wel een geval waarin dat opging:</b></p> <p>.....</p> <p>.....</p> <p>.....</p>	<p><b>Want ik ken wel een geval waarin dat opging:</b></p> <p>.....</p> <p>.....</p> <p>.....</p>
De vraag die u straks wordt gesteld, luidt: waarom is argument A sterker ter ondersteuning van het standpunt dan argument B?	

**Appendix 4B Material used for uncovering laymen criteria for the argument from example: open case with claim 2**

Ik zou u willen vragen om u zich eens in de volgende situatie te verplaatsen. U hebt gesolliciteerd naar een baan als speechschrijver voor het ministerie van Sociale Zaken en Werkgelegenheid. U bent nu op sollicitatiegesprek. Ik ben de manager van de afdeling Personeelszaken en ik moet testen of u goede toespraken kunt schrijven. U krijgt de volgende opdracht:

---

U moet een toespraak schrijven voor de minister. In deze toespraak verdedigt hij de volgende stelling:

**Als werkgevers meer bijdragen aan de kinderopvang, dan zullen werknemers werk en zorg gemakkelijker kunnen combineren.**

Als speechschrijver weet u dat u een standpunt moet ondersteunen met sterke argumenten. Uw taak is nu om twee argumenten te geven die het standpunt ondersteunen. U kiest ervoor om het standpunt te ondersteunen met voorbeelden. Geef

- één sterk voorbeeld dat het standpunt goed ondersteunt
- en
- één zwak voorbeeld dat het standpunt juist niet goed ondersteunt.

Nadat u een sterk en een zwak argument heeft gegeven, krijgt u de vraag waarom het ene argument sterker is ter ondersteuning van het standpunt dan het andere argument.

Het doel van deze opdracht bij uw sollicitatie is erachter te komen of u in staat bent onderscheid te maken tussen sterke en zwakke argumenten.

<b>Als werkgevers meer bijdragen aan de kinderopvang, dan zullen werknemers werk en zorg gemakkelijker kunnen combineren.</b>	
A. Het sterke argument:	B. Het zwakke argument:
<p><b>Want ik ken wel een geval waarin dat opging:</b></p> <p>.....</p> <p>.....</p> <p>.....</p>	<p><b>Want ik ken wel een geval waarin dat opging:</b></p> <p>.....</p> <p>.....</p> <p>.....</p>
De vraag die u straks wordt gesteld, luidt: waarom is argument A sterker ter ondersteuning van het standpunt dan argument B?	

**Appendix 4C Material used for uncovering laymen criteria for the argument from example: closed case with claim 3**

De overheid streeft naar een gezonde samenleving. Niet alleen probeert ze ziekten te voorkomen en vroegtijdig op te sporen met bijvoorbeeld vaccinatieprogramma's en bevolkingsonderzoeken, maar ook probeert ze een gezonde leefstijl te bevorderen. Stel dat u het volgende standpunt moet verdedigen:

**Als je een gezonde leefstijl nastreeft, dan verklein je de kans op ziektes.**

Om dit standpunt te onderbouwen, kunt u een keuze maken uit de onderstaande argumenten. Welke van deze argumenten vindt u sterk ter ondersteuning van dit standpunt? En welke argumenten vindt u juist zwak ter ondersteuning van dit standpunt? Zou u de argumenten willen ordenen, waarbij 1 het beste argument is en 6 het slechtste argument?

- A) Simon de Wit is minder alcohol gaan drinken en heeft nu weer een goed functionerende lever. Bart van Maanen is gestopt met roken en heeft nu minder last van zijn luchtwegen. Sander Schaafstal is meer gaan sporten waardoor hij zich een stuk vitaler voelt. Dus: als je een gezonde leefstijl nastreeft, dan verklein je de kans op ziektes.
- B) Theo van Gogh (regisseur, televisiemaker en columnist) leefde gezond en die had een goede weerstand opgebouwd tegen ziekten. Dus: als je een gezonde levensstijl nastreeft, dan verklein je de kans op ziektes.
- C) Hanneke de Waal heeft haar maag laten verkleinen en nu heeft ze minder last van pijnlijke gewrichten. Dus: als je een gezonde leefstijl nastreeft, dan verklein je de kans op ziektes.
- D) Robert de Graaf is meer water gaan drinken en nu ziet zijn huid er minder vaal uit. Dus: als je een gezonde leefstijl nastreeft, dan verklein je de kans op ziektes.
- E) Robin van Persie volgt een uitgebalanceerd dieet en is daardoor in goede conditie. Dus: als je een gezonde leefstijl nastreeft, dan verklein je de kans op ziektes.
- F) Maria de Graaf is minder vet gaan eten, Edgar Ritmeester eet meer vezels en granen dan vroeger en Frederieke van Loon eet tegenwoordig meer groenten en fruit. Ze hebben nu minder last van kwaaltjes. Dus: als je een gezonde leefstijl nastreeft, dan verklein je de kans op ziektes.

Keuze:

1. \_\_\_\_\_ 2. \_\_\_\_\_ 3. \_\_\_\_\_ 4. \_\_\_\_\_ 5. \_\_\_\_\_ 6. \_\_\_\_\_

**Appendix 4D Material used for uncovering laymen criteria for the argument from example: closed case with claim 4**

Uit prognoses van de overheid blijkt dat de vraag naar bèta's en technici in 2010 veel groter zal zijn dan het aanbod. De overheid stimuleert scholieren dan ook om een exacte studierichting te kiezen. Stel dat u het volgende standpunt moet verdedigen:

**Als je een exacte studierichting kiest, dan verdien je later een hoog salaris.**

Om dit standpunt te onderbouwen, kunt u een keuze maken uit de onderstaande argumenten. Welke van deze argumenten vindt u sterk ter ondersteuning van dit standpunt? En welke argumenten vindt u juist zwak ter ondersteuning van dit standpunt? Zou u de argumenten willen ordenen, waarbij 1 het beste argument is en 6 het slechtste argument?

- A) Geert-Jan van Harmelen heeft tandheelkunde gestudeerd en verdient nu 5500 euro per maand. Bert Linthorst heeft informatica gestudeerd en verdient 3000 euro per maand. Jasper Simons heeft bouwkunde gedaan en heeft nu een inkomen van 4500 euro per maand. Dus: als je een exacte studierichting kiest, dan verdien je later een hoog salaris.
- B) Philip Freriks, presentator bij het NOS-journaal, heeft scheikunde gestudeerd en die staat in de top 10 van rijkste Nederlanders. Dus: als je een exacte studierichting kiest, dan verdien je later een hoog salaris.
- C) Janneke Oorthuys heeft een opleiding psychologie gedaan en ontvangt nu een bovengemiddeld inkomen. Dus: als je een exacte studierichting kiest, dan verdien je later een hoog salaris.
- D) Peter Philips heeft diergeneeskunde gestudeerd en die heeft een leuke baan. Dus: als je een exacte studierichting kiest, dan verdien je later een hoog salaris.
- E) De directeur van Shell heeft natuurkunde gestudeerd en verdient nu miljoenen euro's per jaar. Dus: als je een exacte studierichting kiest, dan verdien je later een hoog salaris.
- F) Erna Bruin heeft geneeskunde gestudeerd, Paul Blok farmaceutische wetenschappen en Renske Atsma biomedische wetenschappen. Ze verdienen nu 4000 euro per maand. Dus: als je een exacte studierichting kiest, dan verdien je later een hoog salaris.

Keuze:

1. \_\_\_\_\_ 2. \_\_\_\_\_ 3. \_\_\_\_\_ 4. \_\_\_\_\_ 5. \_\_\_\_\_ 6. \_\_\_\_\_

## Appendix 6A Pre-test material and results

Table 6A1 The claims pre-tested, mean agreement ratings on a 7-point scale (*SD* between parentheses; ratings from 3.00 to 5.00 in bold) and an indication of the argument types in support of the selected claims in the experiment.

Claim	Rating	In the experiment supported by an argument from...
1. Een cursus arbeidsmarktorientatie leidt tot betere kansen op de arbeidsmarkt voor studenten die afgestudeerd zijn in de psychologie. / A labour market orientation course leads to better chances in the labour market for students with a degree in Psychology.	5.25 (1.30)	
2. Het draaien van Mozarts muziek in woonwinkels leidt tot een hogere omzet. / Playing Mozart's music in home furnishing shops leads to a higher turnover.	<b>4.07</b> <b>(1.35)</b>	Cause to effect
3. Het drinken van appelsap kan ervoor zorgen dat je eerder aan gaatjes in je tanden moet worden behandeld. / Drinking apple juice may cause you to need treatment for dental caries sooner.	5.30 (1.11)	
4. Het toenemend gebruik van mobiele telefoons achter het stuur leidt tot meer verkeersongelukken. / The increased use of mobile phones in traffic leads to more traffic accidents. (+)	6.43 (.71)	
5. Communiceren via e-mail kan de band met een oom of tante in het buitenland versterken. / Communicating through email may strengthen the relationship with an uncle or aunt abroad.	6.17 (.81)	
6. Het eten van gember verhoogt de prestaties bij hardlopen. / Eating ginger increases running performance.	<b>3.77</b> <b>(.73)</b>	Cause to effect
7. Het invoeren van strengere regels op het VMBO kan voorkomen dat jongeren de dupe worden van verbale intimidatie. / Introducing stricter rules in VMBO <sup>107</sup> may prevent adolescents becoming victims of verbal intimidation.	<b>4.87</b> <b>(1.59)</b>	
8. Het geraas van auto's die langs scholen rijden leidt tot slechtere schoolprestaties van leerlingen met ADHD. / The noise of cars driving past schools leads to worsened scholastic results for pupils with ADHD.	5.05 (1.48)	
9. Als kinderen een voorlichtingsfilmpje zien over veilig gebruik van een papierschaar, dan snijden ze zich minder vaak in de vingers. / If children watch an instructional film about the safe use of paper scissors, they will cut their fingers less often.	<b>4.27</b> <b>(1.59)</b>	Cause to effect

<sup>107</sup> VMBO (voorbereidend middelbaar beroepsonderwijs) is an educational level in the Netherlands, literally translated as preparatory middle-level vocational education.



10. Als het Van Gogh museum in Amsterdam meer financiële ondersteuning krijgt, dan nemen de aantallen museumbezoekers uit het buitenland toe. / If the Van Gogh museum in Amsterdam receives more financial support, the amounts of museum visitors from abroad will increase.	<b>4.18</b> <b>(1.57)</b>	Cause to effect
11. Het gebruik van shampoo heeft mogelijk schadelijke effecten op de waterkwaliteit in sloten. / The use of shampoo possibly has detrimental effects on the water quality in ditches.	<b>4.25</b> <b>(1.34)</b>	Cause to effect
12. Het gebruik van insecticiden in land- en tuinbouw kan leiden tot een toename van rupsen. / The use of insecticides in agriculture and horticulture may lead to an increase in caterpillars.	<b>3.70</b> <b>(1.76)</b>	Cause to effect
13. Het regelmatig eten van pizza's verkleint de kans op het krijgen van kanker. / The regular consumption of pizzas reduces the risk of cancer. (-)	2.70 (1.44)	
14. Wanneer jeugdthons beter gecontroleerd worden door de overheid, dan leidt dat tot minder bierverslaving onder jongeren. / If youth clubs are better supervised by the government, this will lead to less beer addiction among adolescents.	<b>4.27</b> <b>(1.89)</b>	Circular argument
15. Het eten van pinda's beschermt tegen galstenen. / Eating peanut butter protects against gallstones.	<b>3.60</b> <b>(.98)</b>	Cause to effect
16. Als immigranten een Nederlands dagblad lezen, dan gaan ze vlotter Nederlands spreken. / Immigrants reading Dutch newspapers will lead to them speaking Dutch more fluently.	5.28 (1.50)	
17. Als sollicitanten zich beter voorbereiden op een sollicitatiegesprek, dan is bij een eventuele salarisonderhandeling de kans op een goed salaris groter. / If applicants prepare better on a job interview, the chance of a good salary increases during a salary negotiation.	5.58 (1.11)	
18. Olijfolie kan pijn bij reumapatiënten bestrijden. / Olive oil may relieve pain in rheumatism patients.	<b>3.45</b> <b>(1.09)</b>	Cause to effect
19. Als je regelmatig Poker speelt, dan leer je beter optellen en aftrekken. / Playing Poker regularly will lead to improved skill in adding and subtracting.	<b>4.12</b> <b>(1.62)</b>	Cause to effect
20. Als de toegangsprijzen voor tentoonstellingen lager worden, dan melden meer studenten zich aan bij beroepsopleidingen voor beeldende kunsten. / If the entrance fees for expositions are reduced, more students will sign up for visual arts degree programmes.	<b>3.15</b> <b>(1.85)</b>	
21. Het gebruik van een lichtgewicht muis en regelmatig ontspannen vermindert de kans op RSI bij mensen die vaak urenlang achter de computer zitten. / The use of a lightweight mouse and regular relaxation reduces the risk of RSI for people who often spend hours sitting at their computers. (+)	5.78 (1.27)	
22. Als de hoeveelheid testosteron bij jongeren wordt verlaagd, dan neigen ze minder naar henneproducten. / If the amount of testosterone in adolescents is reduced they will be less inclined to use cannabis.	2.98 (1.48)	

23. Als er meer voorlichting komt over de werking van koffie en thee, dan zullen minder mensen daaraan verslaafd raken. / If there is more information on the effects of coffee and tea less people will become addicted to them.	<b>3.90</b> <b>(1.72)</b>	Cause to effect
24. Verplichte rijlessen voor mensen boven de 70 jaar kunnen hun angst in het verkeer verminderen. / Obligatory driving lessons for people over 70 may reduce their fear in traffic.	<b>4.37</b> <b>(1.76)</b>	Authority
25. Vertegenwoordigers voelen zich meer betrokken bij hun bedrijf als ze thuis en onderweg gebruik kunnen maken van nieuwe media, zoals een laptop met internetverbinding. / Representatives feel more involved with their companies if they can use new media, such as a laptop with an internet connection, at home or on the way.	<b>4.35</b> <b>(1.69)</b>	Authority
26. Zwangere vrouwen die hun cholesterolgehalte verlagen, lopen minder risico om een te vroeg geboren kind te baren. / Pregnant women who decrease their cholesterol level run a lower risk of having a preterm delivery.	<b>3.80</b> <b>(.88)</b>	Authority
27. Door veel te fietsen over hobbelige, slechte wegen krijg je sterkere botten. / Riding a bike on bumpy, bad roads promotes stronger bones. (-)	2.23 (1.25)	
28. Minder kraamzorg kan leiden tot psychische problemen van de pasgeborene op latere leeftijd. / Less maternity care may lead to psychological problems for the infant at a later age.	<b>3.05</b> <b>(1.50)</b>	
29. Kinderen die op jonge leeftijd Monopoly leren spelen kunnen later beter met geld omgaan. / Children who learn to play Monopoly at a young age are more capable of handling money when they get older.	<b>3.85</b> <b>(1.76)</b>	Authority
30. Het eten van honing leidt tot betere prestaties bij duursporten. / Eating honey leads to better performance in endurance sports.	<b>3.90</b> <b>(1.57)</b>	Authority
31. Hoogrendementsketels leiden tot minder uitstoot van vervuilende gassen. / High efficiency boilers lead to less emissions of polluting gases.	5.00 (1.62)	
32. Een stijging van de postzegeltarieven verhoogt de kwaliteit van de postbezorging. / Raising the price of stamps will increase the quality of mail delivery.	1.97 (1.17)	
33. Als je vaak met het gezin eet in je jeugd, dan eet je op latere leeftijd waarschijnlijk gezonder. / Frequently eating together as a family during childhood, will probably lead to a healthier diet at a later age.	<b>4.68</b> <b>(1.69)</b>	Circular argument
34. Als er minder naar olie wordt geboord in zeeën, dan planten bruinvissen zich eerder voort. / If there is less oil drilling in seas, porpoises will reproduce earlier.	<b>4.02</b> <b>(1.35)</b>	Authority
35. Het lopen op duurdere sportschoenen verlaagt het risico op spierscheuringen. / Walking on more expensive sneakers will lower the risk of muscular ruptures.	<b>3.88</b> <b>(2.00)</b>	Circular argument
36. Als je na een boswandeling je kleren wast op minimaal 60 graden, dan heb je minder kans op een tekenbeet. / If you wash your clothes at at least 60 degree Celsius after a walk	2.28 (1.59)	

in the woods, this will reduce the chance of a tick bite.		
37. De invoering van een voorlopig rijbewijs kan roekeloos rijgedrag van automobilisten die net hun rijbewijs hebben tegengaan. / The introduction of a provisional driving licence may prevent reckless driving by drivers who have just passed their driving test. (+)	5.48 (1.81)	
38. Als je de hygiëne van je voeten verwaarloost, dan kun je eerder malaria krijgen in landen waar die ziekte heerst. / If you neglect foot hygiene, you will get malaria sooner in countries where the risk of malaria is high.	2.48 (1.68)	
39. Met mate rode wijn drinken kan helpen tegen voedselvergiftiging. / Drinking red wine moderately may help against food poisoning.	2.98 (1.46)	
40. Als bromfietrijders gestimuleerd worden om het openbaar vervoer te nemen, dan neemt het aantal dodelijke ongevallen in het verkeer af. / If moped drivers are being stimulated to use public transportation, the number of deadly accidents in traffic will be reduced.	<b>4.85</b> <b>(1.63)</b>	
41. Mensen die lijden aan het chronische vermoeidheidssyndroom ME gaan zich waarschijnlijk beter voelen als ze melkchocolade eten. / People who suffer from the chronic fatigue syndrom ME will probably feel better if they eat milk chocolat.	<b>3.28</b> <b>(1.49)</b>	Authority
42. Groene theebladeren kun je gebruiken ter voorkoming van ontstoken tandvlees. / Green tea leaves can be used to prevent infected gums.	<b>3.75</b> <b>(1.19)</b>	Authority
43. Als vliegtuigen zo lang mogelijk hoog blijven vliegen, dan komt dat de leefomgeving van zoetwatervissen ten goede. / If airplanes fly high as long as possible, the living climate of freshwater fishes will be improved.	2.87 (1.54)	
44. Een landelijke campagne vóór borstvoeding verkleint de kans op roken bij kinderen die geen borstvoeding hebben gehad. / A national campaign in favor of breastfeeding will decrease the chance of smoking for children who have not been breastfed. (-)	2.85 (1.59)	
45. Het kijken naar praatprogramma's op televisie kan helpen om trauma's te verwerken. / Watching talkshows on television may help in processing traumas.	<b>4.43</b> <b>(1.38)</b>	Authority
46. Als Nederlanders met verkooppromoties in supermarkten worden gestimuleerd om ovenfriet te kopen, dan gaan ze minder snel naar fastfoodketens, zoals McDonald's. / If the Dutch are stimulated to buy oven chips by special offers in supermarkets, they will go to fastfood chains like McDonald's less often.	<b>3.20</b> <b>(1.57)</b>	Authority
47. De aanleg van extra voorsorteervakken op kruispunten leidt in grote steden tot een aanzienlijk betere verkeersdoorstroming. / The construction of extra traffic lanes on crossings leads to a substantially better flow of traffic in big cities.	5.00 (1.24)	
48. Aandacht op televisie voor belangrijke wereldkampioenschappen leidt tot een toeloop van	5.68 (1.19)	

jeugdleden bij sportverenigingen. / Paying attention to important world championships on television leads to more youth joining sport clubs.		
49. Ouders die door hun puberkinderen zijn geïnterviewd over hun levensverhaal krijgen een betere relatie met hun zoon of dochter. / Parents who have been interviewed by their adolescent children on their life stories will have a better relationship with their son or daughter.	<b>4.83</b> <b>(1.36)</b>	
50. Als je je kinderen thuis laat blijven bij slecht weer, dan snoepen ze meer. / If you let your children stay home in bad weather conditions, they will eat more sweets.	<b>4.17</b> <b>(1.81)</b>	Example
51. Als mensen een keer per week een denksport beoefenen, dan kunnen ze op school of op het werk beter nadenken. / If people do a mental exercise once a week, this will improve their ability to think at school or work.	5.02 (1.41)	
52. Basisscholen met een goede ventilatie zorgen ervoor dat hun leerlingen beter presteren in schoolvakken. / Elementary schools with good ventilation cause their pupils to perform better in school subjects.	5.80 (1.18)	
53. Als grote Nederlandse dagbladen aandacht besteden aan vermiste personen, dan worden die eerder teruggevonden. / If large Dutch newspapers pay attention to missing persons, they will be found sooner.	5.23 (1.35)	
54. Uitgaansgelegenheden die voorzien zijn van camera's geven de bezoekers een veiliger gevoel. / People feel safer going out in places with camera surveillance than places without. (+)	5.30 (1.49)	
55. Het eten van veel groenten zorgt voor een gezondere bloeddruk. / Eating lots of vegetables leads to a healthier blood pressure.	5.05 (1.30)	
56. Angst om auto te rijden kan worden opgelost door een dag mee te rijden met een vrachtwagenchauffeur. / Fear of driving can be cured by driving along with a truck driver for a day. (-)	<b>3.15</b> <b>(1.37)</b>	
57. Mensen met een evenwichtstoornis hebben baat bij het beoefenen van een balsport. / People with a balance disorder benefit from playing ball sports.	<b>4.60</b> <b>(1.15)</b>	Example
58. Als politiemedewerkers op belangrijke feestdagen actie voeren, dan gaat dat ten koste van de reputatie van de politie. / If police officers protest during important holidays, this will damage the reputation of the police.	<b>4.35</b> <b>(2.11)</b>	Circular argument
59. Als bejaarden minder medicijnen slikken, dan lopen ze minder risico om te vallen. / If elderly people take less medication, they will run less risk of falling.	<b>3.50</b> <b>(1.90)</b>	Example
60. Als meer mensen vogelkastjes in hun tuin maken voor zangvogels, dan zijn de nesten van die vogels beter beschermd. / If more people make birdhouses in their garden for songbirds, the nests of these birds will be better protected.	<b>4.55</b> <b>(1.87)</b>	Example
61. Mensen die dagelijks een slaginstrument bespelen, lopen een groter risico om doofheid voor hoge tonen te	<b>4.80</b> <b>(1.59)</b>	Example

ontwikkelen. / People who play a percussion instrument each day run a higher risk of developing deafness to high tones.		
62. Als je een kamerplant op je slaapkamer houdt, dan kan dat leiden tot een betere nachtrust. / If you have an indoor plant in your bedroom, this may lead to a better night's sleep.	<b>3.68</b> <b>(1.93)</b>	Example
63. Als kinderen op de kinderboerderij dieren aaien, dan krijgen ze eerder last van huidproblemen. / If children pet animals at the children's farm, this will lead to a higher risk of skin problems.	2.40 (1.52)	
64. Televisiereclame voor het gebruik van een openbaar vervoermiddel leidt ertoe dat meer mensen met dat vervoermiddel gaan reizen. / Television advertising for the use of a particular mode of public transport leads to more people using that mode of transport.	<b>4.20</b> <b>(1.68)</b>	Example
65. Tweelingen die apart van elkaar opgroeien hebben vaker een identieke levensloop dan tweelingen die in hetzelfde gezin opgroeien. / Twins that do not grow up in the same family more frequently have a similar life course than twins that do grow up in the same family. (-)	<b>3.10</b> <b>(1.57)</b>	
66. Als je badtextiel naspoelt met schoonmaakazijn, dan heb je minder kans op huidziekten. / Re-rinsing bath textile with cleaning vinegar will reduce the risk of skin diseases.	<b>3.40</b> <b>(1.45)</b>	Example
67. Het gebruik van een ontspanningsruimte op het werk brengt het ziekteverzuim omlaag. / The use of a relaxation room at work reduces absence at the workplace as a result of illness. (+)	5.63 (1.21)	
68. Het eten van zuivelproducten verlaagt de kans op nachtblindheid. / Eating dairy products reduces the chance of night blindness.	<b>3.25</b> <b>(1.30)</b>	Example
69. Echtgenoten die de naaste familieleden van hun man of vrouw af en toe een kaartje sturen, zullen aardiger worden gevonden door die familieleden. / Partners who send their husband or wife's close relatives a postcard once in a while, will be seen as more sympathetic by these family members.	5.18 (1.57)	
70. Als scholieren les krijgen in klassieke talen, dan is de kans groter dat ze later geschiedenis gaan studeren. / If pupils are schooled in the classical languages, increases the chance that they will go on to study History.	<b>4.22</b> <b>(1.70)</b>	Example

(+) : scored highly probable in Hornikx (2005a)

(-) : scored highly improbable in Hornikx (2005a)

## Appendix 6B Experimental material

### Arguments from authority:

#### Text 1

I	Volgens dr. Emiel Bentink, universitair docent Psychologie aan de Universiteit Utrecht en gepromoveerd op angststoornissen bij ouderen, zijn 70-plussers minder angstig op de weg als ze verplicht worden om rijlessen te nemen.
II	Volgens Emiel Bentink, een tweedejaars student Psychologie die zojuist een college over angststoornissen bij ouderen heeft gevolgd, zijn 70-plussers minder angstig op de weg als ze verplicht worden om rijlessen te nemen.
III	Volgens dr. Emiel Bentink, universitair docent Psychologie aan de Universiteit Utrecht en gepromoveerd op angststoornissen bij jongeren, zijn 70-plussers minder angstig op de weg als ze verplicht worden om rijlessen te nemen.
IV	Volgens dr. Emiel Bentink, stafmedewerker van de Vereniging van Autorijschoolhouders en gepromoveerd op angststoornissen bij ouderen, zijn 70-plussers minder angstig op de weg als ze verplicht worden om rijlessen te nemen.
V	Volgens dr. Emiel Bentink, universitair docent Psychologie aan de Universiteit Utrecht en gepromoveerd op angststoornissen bij ouderen, zijn 70-plussers wat minder onzeker op de weg als ze bereid zijn om af en toe rijles te nemen.
Claim	Verplichte rijlessen voor mensen boven de 70 jaar kunnen hun angst in het verkeer dus verminderen.

#### Text 2

I	Dr. Masha van Rijn, gepromoveerd op nieuwe media en bedrijfskunde en nu in dienst van het Ministerie van Economische Zaken, stelt dat vertegenwoordigers een hogere betrokkenheid bij hun bedrijf vertonen wanneer ze nieuwe media kunnen gebruiken.
II	Masha van Rijn, studente aan de HBO-opleiding Nieuwe Media en Bedrijfskunde aan de Hogeschool INHOLLAND, stelt dat vertegenwoordigers een hogere betrokkenheid bij hun bedrijf vertonen wanneer ze nieuwe media kunnen gebruiken.
III	Dr. Masha van Rijn, gepromoveerd op recht en economie en nu in dienst van het Ministerie van Economische Zaken, stelt dat vertegenwoordigers een hogere betrokkenheid bij hun bedrijf vertonen wanneer ze nieuwe media kunnen gebruiken.
IV	Dr. Masha van Rijn, gepromoveerd op nieuwe media en bedrijfskunde en nu in dienst van internetleverancier UPC, stelt dat vertegenwoordigers een hogere betrokkenheid bij hun bedrijf vertonen wanneer ze nieuwe media kunnen gebruiken.
V	Dr. Masha van Rijn, gepromoveerd op het gebied van nieuwe media en bedrijfskunde en nu in dienst van het Ministerie van Economische Zaken, stelt dat vertegenwoordigers wat intensiever bezig zijn met hun werkzaamheden wanneer ze nieuwe media kunnen gebruiken.
Claim	Dus: vertegenwoordigers voelen zich meer betrokken bij hun bedrijf als ze thuis en onderweg gebruik kunnen maken van nieuwe media, zoals een

	laptop met internetverbinding.
--	--------------------------------

### Text 3

I	Dr. Karin Heimnitz, gepromoveerd op problemen bij pasgeboren kinderen, zegt dat zwangere vrouwen die hun cholesterolgehalte verlagen minder kans hebben op vroeggeboorte.
II	Karin Heimnitz, student Verloskunde aan de Hogeschool Utrecht, zegt dat zwangere vrouwen die hun cholesterolgehalte verlagen minder kans hebben op vroeggeboorte.
III	Dr. Karin Heimnitz, gepromoveerd op aandoeningen van het zenuwstelsel, zegt dat zwangere vrouwen die hun cholesterolgehalte verlagen minder kans hebben op vroeggeboorte.
IV	Dr. Karin Heimnitz, gepromoveerd op problemen bij pasgeboren kinderen en nu werkzaam bij het Bechel Institute, zegt dat zwangere vrouwen die hun cholesterolgehalte verlagen minder kans hebben op vroeggeboorte.
V	Dr. Karin Heimnitz, gepromoveerd op problemen bij pasgeboren kinderen, zegt dat zwangere vrouwen die hun cholesterolgehalte in het laatste stadium van de zwangerschap onder controle houden minder kans hebben op vroeggeboorte.
Claim	Zwangere vrouwen die hun cholesterolgehalte verlagen, lopen dus minder risico om een te vroeg geboren kind te baren.

### Text 4

I	Op de website van het vakblad Pedagogiek in de Praktijk beweren pedagogen dat kinderen die vroeg Monopoly leren spelen later beter met geld omgaan.
II	Op het discussieforum van het elektronisch magazine Ouders Online beweren bezoekers dat kinderen die vroeg Monopoly leren spelen later beter met geld omgaan.
III	Op de website van het vakblad Geschiedenis van de Onderwijskunde beweren onderwijskundigen dat kinderen die vroeg Monopoly leren spelen later beter met geld omgaan.
IV	Op de website "Speelwijzer" van de Verenigde Speelgoedfabrikanten beweren pedagogen dat kinderen die vroeg Monopoly leren spelen later beter met geld omgaan.
V	Op de website van het vakblad Pedagogiek in de Praktijk beweren pedagogen dat kinderen die onder begeleiding van ouders en leraren Monopoly leren spelen eerder de waarde van geld leren kennen.
Claim	Kinderen die op jonge leeftijd Monopoly leren spelen kunnen later dus beter met geld omgaan.

### Text 5

I	Onderzoekers van de afdeling Bewegingswetenschappen van de Vrije Universiteit Amsterdam zijn van mening dat het gebruik van honing de prestatie tijdens duurtraining verbetert.
II	Leerlingen van het Holland College, opleiding MBO-Voeding en Sport, zijn van mening dat het gebruik van honing de prestatie tijdens duurtraining verbetert.
III	Onderzoekers van de afdeling Revalidatiegeneeskunde van de Vrije Universiteit Amsterdam zijn van mening dat het gebruik van honing de prestatie tijdens duurtraining verbetert.

IV	Onderzoekers van de Vrije Universiteit die samenwerken met Langnese Honing, de Duitse honingproducent, zijn van mening dat het gebruik van honing de prestatie tijdens duurtraining verbetert.
V	Onderzoekers van de afdeling Bewegingswetenschappen van de Vrije Universiteit Amsterdam zijn van mening dat bepaalde voedingsstoffen, die ook wel in honing zitten, sportprestaties kunnen verbeteren
Claim	Het eten van honing leidt dus tot betere prestaties bij duursporten.

#### Text 6

I	Volgens onderzoekers van de Universiteit Antwerpen, departement Biologie, zullen bruinvissen zich eerder voortplanten, wanneer het boren naar olie aan banden wordt gelegd.
II	Volgens leden van de Vlaamse Werkgroep Biologie zullen bruinvissen zich eerder voortplanten, wanneer het boren naar olie aan banden wordt gelegd.
III	Volgens onderzoekers van de Universiteit Antwerpen, departement Natuurkunde, zullen bruinvissen zich eerder voortplanten, wanneer het boren naar olie aan banden wordt gelegd.
IV	Volgens onderzoekers van de Vlaamse tak van Save Our Oceans zullen bruinvissen zich eerder voortplanten, wanneer het boren naar olie aan banden wordt gelegd.
V	Volgens onderzoekers van de Universiteit Antwerpen, departement Biologie, zal het leefklimaat van bruinvissen aanzienlijk aan kwaliteit winnen, wanneer het boren naar olie aan banden wordt gelegd.
Claim	Als er dus minder naar olie wordt geboord in zeeën, dan planten bruinvissen zich eerder voort.

#### Text 7

I	ME-specialisten in het Universitair Medisch Centrum Utrecht beweren dat melkchocolade bevorderlijk is voor het welbevinden van mensen die aan ME lijden.
II	Co-assistenten die in het Universitair Medisch Centrum Utrecht met ME-patiënten werken, beweren dat melkchocolade bevorderlijk is voor het welbevinden van mensen die aan ME lijden.
III	ADHD-specialisten in het Universitair Medisch Centrum Utrecht beweren dat melkchocolade bevorderlijk is voor het welbevinden van mensen die aan ME lijden.
IV	ME-specialisten in het Universitair Medisch Centrum beweren op de website van Milka dat melkchocolade bevorderlijk is voor het welbevinden van mensen die aan ME lijden.
V	ME-specialisten in het Universitair Medisch Centrum Utrecht vermoeden dat bestanddelen die ook wel in melkchocolade worden verwerkt, bevorderlijk kunnen zijn voor het welbevinden van mensen die aan ME lijden.
Claim	Dus: mensen die lijden aan het chronische vermoeidheidssyndroom ME gaan zich waarschijnlijk beter voelen als ze melkchocolade eten.

#### Text 8

I	In een recente publicatie in het vakblad voor tandheelkundigen worden groene theebladeren geadviseerd, omdat ze beschermen tegen ontstoken tandvlees.
---	---



II	In een recente nieuwsbrief van de Patiëntenbelangenvereniging Tandheelkunde worden groene theebladeren geadviseerd, omdat ze beschermen tegen ontstoken tandvlees.
III	In een recente publicatie in het vakblad voor erfelijkheidsdeskundigen worden groene theebladeren geadviseerd, omdat ze beschermen tegen ontstoken tandvlees.
IV	In een recente publicatie op de website van Pickwick worden door tandheelkundigen groene theebladeren geadviseerd, omdat ze beschermen tegen ontstoken tandvlees.
V	In een recente publicatie in het vakblad voor tandheelkundigen worden groene theebladeren geadviseerd, omdat ze, als ze voldoende vers zijn, beschermen tegen ontstoken tandvlees.
Claim	Groene theebladeren kun je dus gebruiken ter voorkoming van ontstoken tandvlees.

#### Text 9

I	In een rapport van media-onderzoekers van de Universiteit van Amsterdam staat dat slachtoffers van ongevallen en geweld gebaat zijn bij het kijken naar praatprogramma's op televisie.
II	In een werkstuk van studenten van de HBO-opleiding Media en Samenleving staat dat slachtoffers van ongevallen en van geweld gebaat zijn bij het kijken naar praatprogramma's op televisie.
III	In een rapport van verslavingsonderzoekers van de Universiteit van Amsterdam staat dat slachtoffers van ongevallen en geweld gebaat zijn bij het kijken naar praatprogramma's op televisie.
IV	In een rapport van media-onderzoekers van de Universiteit van Amsterdam, geschreven in opdracht van Endemol, staat dat slachtoffers van ongevallen en geweld gebaat zijn bij het kijken naar praatprogramma's op televisie.
V	In een rapport van media-onderzoekers van de Universiteit van Amsterdam staat dat slachtoffers van ongevallen en geweld gebaat zijn bij het kijken naar praatprogramma's op televisie, als dat gecombineerd wordt met andere vormen van therapie.
Claim	Het kijken naar praatprogramma's op televisie kan dus helpen om trauma's te verwerken.

#### Text 10

I	Volgens marketingprofessionals op Marketing Online, de website van het Tijdschrift voor Marketing, zullen minder Nederlanders fastfoodketens bezoeken wanneer ze in supermarkten aangespoord worden om ovenfriet aan te schaffen.
II	Volgens Xanders weblog, een weblog over marketing, zullen minder Nederlanders fastfoodketens bezoeken wanneer ze in supermarkten aangespoord worden om ovenfriet aan te schaffen.
III	Volgens IT-professionals op Automatisering Online, de website van het Tijdschrift voor Automatisering, zullen minder Nederlanders fastfoodketens bezoeken wanneer ze in supermarkten aangespoord worden om ovenfriet aan te schaffen.
IV	Volgens marketingprofessionals op de website van Aviko, producent van friet en andere aardappelproducten, zullen minder Nederlanders

	fastfoodketens bezoeken wanneer ze in supermarkten aangespoord worden om ovenfriet aan te schaffen.
V	Volgens Marketing Online, de website van het Tijdschrift voor Marketing, zullen minder Nederlanders geneigd zijn om friet te bestellen bij fastfoodketens wanneer ze in supermarkten aangespoord worden om ovenfriet aan te schaffen.
Claim	Dus: als Nederlanders met verkooppromoties in supermarkten worden gestimuleerd om ovenfriet te kopen, dan gaan ze minder snel naar fastfoodketens, zoals McDonald's.

**Argument from cause to effect:**

**Text 1**

VI	Als woonwinkels Mozart draaien, dan heeft dat een positief effect op de stemming van de consumenten. Dat verhoogt de omzet.
VII	Als woonwinkels Mozart draaien, dan blijven consumenten daar langer naar de muziek luisteren. Dat verhoogt de omzet.
VIII	In de meeste gevallen behalen woonwinkels financieel voordeel wanneer ze klassieke muziek spelen.
IX	In de meeste gevallen behalen woonwinkels financieel voordeel wanneer ze populaire muziek spelen.
X	In de meeste gevallen behalen woonwinkels reclamevoordeel wanneer ze klassieke muziek spelen.
Claim	Het draaien van Mozarts muziek in woonwinkels leidt dus tot een hogere omzet.

**Text 2**

VI	Als je gember eet, dan worden je spijsvertering en bloedcirculatie gestimuleerd. Dat leidt tot betere hardloopprestaties.
VII	Als je gember eet, dan wordt je lichaamstemperatuur verhoogd. Dat leidt tot betere hardloopprestaties.
VIII	Het eten van specerijen leidt over het algemeen tot betere prestaties bij duursporten.
IX	Het eten van groenten leidt over het algemeen tot betere prestaties bij duursporten.
X	Het eten van specerijen leidt over het algemeen tot betere prestaties bij krachtsporten.
Claim	Het eten van gember verhoogt dus de prestaties bij hardlopen.

**Text 3**

VI	Kinderen een voorlichtingsfilmpje laten zien over het veilig gebruik van een papierschaar zorgt ervoor dat ze een schaar op een veilige manier weten te gebruiken. Ze zullen zich dan minder gauw snijden.
VII	Kinderen een voorlichtingsfilmpje laten zien over het veilig gebruik van een papierschaar zorgt ervoor dat ze minder snel een schaar durven gebruiken. Ze zullen zich dan minder gauw snijden.
VIII	Voorlichtingsfilms over het veilig gebruik van handgereedschap zorgen meestal voor een afname van verwondingen.
IX	Voorlichtingsfilms over het veilig gebruik van machinereedschap zorgen

	meestal voor een afname van verwondingen.
X	Voorlichtingsfilms over het veilig gebruik van handgereedschap zorgen meestal voor een afname van kneuzingen.
Claim	Dus: als kinderen een voorlichtingsfilmpje zien over het veilig gebruik van een papierschaar, dan snijden ze zich minder vaak in de vingers.

#### Text 4

VI	Met meer financiële ondersteuning kan het Van Gogh museum in Amsterdam meer geld steken in promotie-activiteiten. De aantallen buitenlandse museumbezoekers zullen daardoor stijgen.
VII	Met meer financiële ondersteuning kan het Van Gogh museum in Amsterdam meer geld steken in sanitaire voorzieningen. De aantallen buitenlandse museumbezoekers zullen daardoor stijgen.
VIII	Het financieel steunen van kunstmusea leidt doorgaans tot een toename van museumtoerisme.
IX	Het financieel steunen van oudheidkundige musea leidt doorgaans tot een toename van museumtoerisme.
X	Het financieel steunen van kunstmusea leidt doorgaans tot een toename van massatoerisme.
Claim	Dus: als het Van Gogh museum meer financiële ondersteuning krijgt, dan nemen de aantallen museumbezoekers uit het buitenland toe.

#### Text 5

VI	Als je shampoo gebruikt, dan komen milieubelastende stoffen in de afvoer terecht. Dat kan een schadelijk effect hebben op de waterkwaliteit in sloten.
VII	Als je shampoo gebruikt, dan blijf je wat langer onder de douche staan. Dat kan een schadelijk effect hebben op de waterkwaliteit in sloten.
VIII	Cosmetica die het lichaam reinigen hebben doorgaans een negatieve invloed op het oppervlaktewater.
IX	Cosmetica die het lichaam beschermen hebben doorgaans een negatieve invloed op het oppervlaktewater.
X	Cosmetica die het lichaam reinigen hebben doorgaans een negatieve invloed op het grondwater.
Claim	Het gebruik van shampoo heeft dus mogelijk schadelijke effecten op de waterkwaliteit in sloten.

#### Text 6

VI	Als men insecticiden gebruikt, dan kunnen rupsen er op den duur resistent tegen worden. En dan wordt een rupsenplaag alleen maar erger.
VII	Als men insecticiden gebruikt, dan kan het aantal kevers op den duur afnemen. En dan wordt een rupsenplaag alleen maar erger.
VIII	Het gebruik van chemische bestrijdingsmiddelen zorgt meestal voor een toename van plaaginsecten.
IX	Het gebruik van biologische bestrijdingsmiddelen zorgt meestal voor een toename van plaaginsecten.
X	Het gebruik van chemische bestrijdingsmiddelen zorgt meestal voor een toename van stekende insecten.
Claim	Het gebruik van insecticiden in land- en tuinbouw kan dus leiden tot een toename van rupsen.

**Text 7**

VI	Pindakaas verlaagt het cholesterolgehalte in het bloed en vermindert daarmee de kans op galstenen.
VII	Pindakaas vermindert het hongergevoel en verlaagt daarmee de kans op galstenen.
VIII	Broodbeleg van plantaardige oorsprong helpt doorgaans tegen galblaasaandoeningen.
IX	Broodbeleg van dierlijke oorsprong helpt doorgaans tegen galblaasaandoeningen.
X	Broodbeleg van plantaardige oorsprong helpt doorgaans tegen lymfe-aandoeningen
Claim	Het eten van pindakaas beschermt dus tegen galstenen.

**Text 8**

VI	Inname van olijfolie leidt ertoe dat je ontstekingsremmende stoffen binnenkrijgt. Dat kan een pijnstillende werking hebben.
VII	Inname van olijfolie leidt ertoe dat je meer vocht binnenkrijgt. Dat kan een pijnstillende werking hebben.
VIII	Plantaardige oliën helpen meestal bij gewrichtsaandoeningen.
IX	Dierlijke oliën helpen meestal bij gewrichtsaandoeningen.
X	Plantaardige oliën helpen meestal bij zenuwaandoeningen
Claim	Dus: olijfolie kan pijn bij reumapatiënten bestrijden.

**Text 9**

VI	Bij poker kun je spelen met fiches die een geldwaarde voorstellen. Dan leer je vanzelf beter optellen en aftrekken.
VII	Bij poker kun je bluffen over de kwaliteit van je speelkaarten. Dan leer je vanzelf beter optellen en aftrekken.
VIII	Van het spelen van kaartspellen word je doorgaans beter in rekenen.
IX	Van het spelen van denkspellen word je doorgaans beter in rekenen.
X	Van het spelen van kaartspellen word je doorgaans beter in schatten.
Claim	Dus: als je regelmatig poker speelt, dan leer je beter optellen en aftrekken.

**Text 10**

VI	Meer voorlichting over koffie en thee zorgt ervoor dat mensen een afwijzende houding ten aanzien van deze producten krijgen. Dat minder mensen er dan verslaafd aan raken, is het gevolg.
VII	Meer voorlichting over koffie en thee zorgt ervoor dat bedrijven en kantoren minder koffieautomaten neerzetten. Dat minder mensen er dan verslaafd aan raken, is het gevolg.
VIII	Meestal geldt: hoe meer voorlichting over stimulerende middelen, des te minder verslaving aan genotmiddelen.
IX	Meestal geldt: hoe meer voorlichting over verdovende middelen, des te minder verslaving aan genotmiddelen.
X	Meestal geldt: hoe meer voorlichting over stimulerende middelen, des te minder verslaving aan geneesmiddelen
Claim	Dus: als er meer voorlichting komt over de werking van koffie en thee, dan zullen minder mensen daaraan verslaafd raken.

### Arguments from example:

#### Text 1

XI	De 13-jarige Gertjan Vriesering greep vaker naar de chocolade, toen het weer slecht was en hij binnen moest blijven.
XII	De 13-jarige Gertjan Vriesering greep vaker naar de chips, toen het weer slecht was en hij binnen moest blijven.
XIII	De 13-jarige Gertjan Vriesering greep vaker naar de chocolade, toen het weer slecht was en hij binnen moest blijven. Of neem de 10-jarige Emilie Kruithof. Toen ze van haar moeder thuis moest blijven vanwege slecht weer, begon ze veel drop te eten.
XIV	De 13-jarige Gertjan Vriesering greep vaker naar de chocolade, toen het weer slecht was en hij binnen moest blijven. Of neem de 10-jarige Emilie Kruithof. Toen ze van haar moeder thuis moest blijven vanwege slecht weer, begon ze veel chips te eten.
XV	De 13-jarige Gertjan Vriesering greep vaker naar de chips, toen het weer slecht was en hij binnen moest blijven. Of neem de 10-jarige Emilie Kruithof. Toen ze van haar moeder thuis moest blijven vanwege slecht weer, begon ze veel pepermint te eten.
Claim	Als je je kinderen thuis laat blijven bij slecht weer, dan snoepen ze dus meer.

#### Text 2

XI	Als mensen die problemen hebben met hun evenwicht regelmatig handbal spelen, dan kunnen ze sneller herstellen.
XII	Als mensen die problemen hebben met hun evenwicht regelmatig waterpolo spelen, dan kunnen ze sneller herstellen.
XIII	Als mensen die problemen hebben met hun evenwicht regelmatig handbal spelen, dan kunnen ze sneller herstellen. Of basketbal: dat helpt ook het evenwichtsgevoel te ontwikkelen.
XIV	Als mensen die problemen hebben met hun evenwicht regelmatig handbal spelen, dan kunnen ze sneller herstellen. Of waterpolo: dat helpt ook het evenwichtsgevoel te ontwikkelen.
XV	Als mensen die problemen hebben met hun evenwicht regelmatig waterpolo spelen, dan kunnen ze sneller herstellen. Of ijshockey: dat helpt ook het evenwichtsgevoel te ontwikkelen.
Claim	Mensen met een evenwichtstoornis hebben dus baat bij het beoefenen van een balsport.

#### Text 3

XI	Toen de 70-jarige Bert Tiemes op advies van de huisarts minder medicijnen innam, maakte hij minder valpartijen.
XII	Toen de 60-jarige Bert Tiemes op advies van de huisarts minder medicijnen innam, maakte hij minder valpartijen.
XIII	Toen de 70-jarige Bert Tiemes op advies van de huisarts minder medicijnen innam, maakte hij minder valpartijen. Of neem de 75-jarige Frans Lagerman. Toen hij stopte met bepaalde medicijnen, viel hij minder vaak.
XIV	Toen de 70-jarige Bert Tiemes op advies van de huisarts minder medicijnen innam, maakte hij minder valpartijen. Of neem de 60-jarige Frans Lagerman. Toen hij stopte met bepaalde medicijnen, viel hij minder

	vaak.
XV	Toen de 60-jarige Bert Tiemes op advies van de huisarts minder medicijnen innam, maakte hij minder valpartijen. Of neem de 63-jarige Frans Lagerman. Toen hij stopte met bepaalde medicijnen, viel hij minder vaak.
Claim	Als bejaarden minder medicijnen slikken, dan lopen ze dus minder risico om te vallen.

#### Text 4

XI	Als je een vogelkastje maakt voor nachtegalen, dan zijn hun eieren beter beschermd.
XII	Als je een vogelkastje maakt voor huismussen, dan zijn hun eieren beter beschermd.
XIII	Als je een vogelkastje maakt voor nachtegalen, dan zijn hun eieren beter beschermd. Ook zanglijsters broeden veiliger in je tuin.
XIV	Als je een vogelkastje maakt voor nachtegalen, dan zijn hun eieren beter beschermd. Ook huismussen broeden veiliger in je tuin.
XV	Als je een vogelkastje maakt voor huismussen, dan zijn hun eieren beter beschermd. Ook koolmezen broeden veiliger in je tuin.
Claim	Dus: als meer mensen vogelkastjes in hun tuin maken voor zangvogels, dan zijn de nesten van die vogels beter beschermd.

#### Text 5

XI	Mensen die iedere dag op een grote trom spelen, lopen een verhoogd risico op gehoorverlies bij hoge tonen.
XII	Mensen die iedere dag op een tamboerijn spelen, lopen een verhoogd risico op gehoorverlies bij hoge tonen.
XIII	Mensen die iedere dag op een grote trom spelen, lopen een verhoogd risico op gehoorverlies bij hoge tonen. Dat geldt trouwens ook voor diegenen die dagelijks gong spelen.
XIV	Mensen die iedere dag op een grote trom spelen, lopen een verhoogd risico op gehoorverlies bij hoge tonen. Dat geldt trouwens ook voor diegenen die dagelijks tamboerijn spelen.
XV	Mensen die iedere dag op een tamboerijn spelen, lopen een verhoogd risico op gehoorverlies bij hoge tonen. Dat geldt trouwens ook voor diegenen die dagelijks xylofoon spelen.
Claim	Mensen die dus dagelijks een slaginstrument bespelen, lopen een groter risico om doofheid voor hoge tonen te ontwikkelen.

#### Text 6

XI	Als je een geranium in je slaapkamer zet, dan zou je nachtrust daarbij gebaat kunnen zijn.
XII	Als je bamboe in je slaapkamer zet, dan zou je nachtrust daarbij gebaat kunnen zijn.
XIII	Als je een geranium in je slaapkamer zet, dan zou je nachtrust daarbij gebaat kunnen zijn. Ook een ficus wil nog wel eens bijdragen aan een goede nachtrust.
XIV	Als je een geranium in je slaapkamer zet, dan zou je nachtrust daarbij gebaat kunnen zijn. Ook bamboe wil nog wel eens bijdragen aan een goede nachtrust.
XV	Als je bamboe in je slaapkamer zet, dan zou je nachtrust daarbij gebaat

	kunnen zijn. Ook een cactus wil nog wel eens bijdragen aan een goede nachtrust.
Claim	Als je dus een kamerplant op je slaapkamer houdt, dan kan dat leiden tot een betere nachtrust.

#### Text 7

XI	Mensen gaan eerder met de metro als daarvoor reclame wordt gemaakt op TV.
XII	Mensen gaan eerder met de treintaxi als daarvoor reclame wordt gemaakt op TV.
XIII	Mensen gaan eerder met de metro als daarvoor reclame wordt gemaakt op TV. Of ze pakken eerder de tram, als daarover TV-spotjes worden uitgezonden.
XIV	Mensen gaan eerder met de metro als daarvoor reclame wordt gemaakt op TV. Of ze pakken eerder de treintaxi, als daarover TV-spotjes worden uitgezonden.
XV	Mensen gaan eerder met de treintaxi als daarvoor reclame wordt gemaakt op TV. Of ze pakken eerder de veerpont, als daarover TV-spotjes worden uitgezonden.
Claim	Televisiereclame voor het gebruik van een openbaar vervoermiddel leidt er dus toe dat meer mensen met dat vervoermiddel gaan reizen.

#### Text 8

XI	Het naspoelen van handdoeken met schoonmaakazijn verkleint de kans op aandoeningen van de huid.
XII	Het naspoelen van strandlakens met schoonmaakazijn verkleint de kans op aandoeningen van de huid.
XIII	Het naspoelen van handdoeken met schoonmaakazijn verkleint de kans op aandoeningen van de huid. Ook het naspoelen van washandjes kan de kans op huidaandoeningen verkleinen.
XIV	Het naspoelen van handdoeken met schoonmaakazijn verkleint de kans op aandoeningen van de huid. Ook het naspoelen van strandlakens kan de kans op huidaandoeningen verkleinen.
XV	Het naspoelen van strandlakens met schoonmaakazijn verkleint de kans op aandoeningen van de huid. Ook het naspoelen van badstoffen slippers kan de kans op huidaandoeningen verkleinen.
Claim	Als je badtextiel dus naspoelt met schoonmaakazijn, dan heb je minder kans op huidziekten.

#### Text 9

XI	Het eten van kaas kan nachtblindheid voorkomen.
XII	Het eten van pudding kan nachtblindheid voorkomen.
XIII	Het eten van kaas kan nachtblindheid voorkomen. Ook echte boter kan het risico op deze aandoening verlagen.
XIV	Het eten van kaas kan nachtblindheid voorkomen. Ook pudding kan het risico op deze aandoening verlagen.
XV	Het eten van pudding kan nachtblindheid voorkomen. Ook pap kan het risico op deze aandoening verlagen.
Claim	Het eten van zuivelproducten verlaagt dus de kans op nachtblindheid.

**Text 10**

XI	Alexander kreeg als scholier onderwijs in Oud-Grieks en is geschiedenis gaan studeren.
XII	Alexander kreeg als scholier onderwijs in Hebreeuws en is geschiedenis gaan studeren.
XIII	Alexander kreeg als scholier onderwijs in Oud-Grieks en is geschiedenis gaan studeren. Of neem Stan. Hij kreeg lessen Latijn op de middelbare school en nu is hij begonnen aan een studie Geschiedenis.
XIV	Alexander kreeg als kind onderwijs in Oud-Grieks en is geschiedenis gaan studeren. Of neem Stan. Hij kreeg lessen Hebreeuws op de middelbare school en nu is hij begonnen aan een studie Geschiedenis.
XV	Alexander kreeg als kind onderwijs in Hebreeuws en is geschiedenis gaan studeren. Of neem Stan. Hij kreeg lessen Arabisch op de middelbare school en nu is hij begonnen aan een studie Geschiedenis.
Claim	Als scholieren dus les krijgen in klassieke talen, dan is de kans groter dat ze later geschiedenis gaan studeren.

**Circular arguments:****Text 1**

Een overheid die clubhuizen voor jongeren beter in de gaten houdt, voorkomt dat de jeugd geestelijk en lichamelijk afhankelijk wordt van bier.
Wanneer jeugdthons beter gecontroleerd worden door de overheid, dan leidt dat dus tot minder bierverslaving onder jongeren.

**Text 2**

Wie als kind vele malen in gezinsverband eet, heeft als volwassene mogelijk een gezonder voedingspatroon.
Als je dus vaak met het gezin eet in je jeugd, dan eet je op latere leeftijd waarschijnlijk gezonder.

**Text 3**

Gescheurde spiervezels kun je voorkomen door te rennen op sportschoenen die hoger zijn van prijs.
Het lopen op duurdere sportschoenen verlaagt dus het risico op spierscheuringen.

**Text 4**

Politieagenten die staken of demonstreren op gedenkdagen van grote betekenis, bezorgen hun werkgever een slechtere naam.
Dus: als politiemedewerkers op belangrijke feestdagen actie voeren, dan gaat dat ten koste van de reputatie van de politie.



### Appendix 6C Results manipulation checks

Table 6C1 Mean scores on manipulation checks for arguments from authority (*SD* between parentheses; text numbers correspond to texts numbers in appendix 6B).

Text	Relevant expertise (1=very incompetent in the field the statement is in; 7= very competent in the field the statement is in)			Bias (1=very unbiased towards the field the statement is in; 7=very biased towards the field the statements is in)		
	Strong condition (I)	Weak condition (III)  violates relevant expertise criterion	Test result (all p-values one-tailed)	Strong condition (I)	Weak condition (IV)  violates trustworthiness criterion	Test result (all p-values one-tailed)
1	5.51 (1.52)	4.13 (1.87)	$t(76) = 3.59, p < .001^{***}$	3.89 (1.84)	5.28 (1.50)	$t(75) = -3.63, p < .001^{***}$
2	5.44 (1.14)	4.13 (1.41)	$t(67) = 4.24, p < .001^{***}$	4.31 (1.75)	5.42 (1.42)	$t(75) = -3.06, p < .002^{**}$
3	5.55 (1.03)	4.51 (1.52)	$t(64.4) = 3.39, p < .001^{***}$	3.29 (1.55)	5.26 (1.76)	$t(68) = -4.89, p < .001^{***}$
4	5.05 (1.49)	4.28 (1.57)	$t(76) = 2.22, p < .02^*$	3.79 (1.47)	5.58 (1.46)	$t(68) = -5.07, p < .001^{***}$
5	5.21 (1.15)	4.87 (1.58)	$t(76) = 1.07, p = .145$	3.13 (1.52)	5.13 (1.61)	$t(76) = -5.64, p < .001^{***}$
6	5.49 (1.17)	5.03 (1.58)	$t(76) = 1.47, p = .074$	4.10 (1.60)	5.58 (1.50)	$t(75) = -4.17, p < .001^{***}$
7	5.67 (1.58)	4.06 (1.79)	$t(68) = 3.98, p < .001^{***}$	2.97 (2.01)	4.36 (1.50)	$t(68.4) = -3.43, p < .001^{***}$
8	5.87 (.92)	3.74 (1.70)	$t(68) = 6.28, p < .001^{***}$	2.87 (1.59)	5.26 (1.62)	$t(68) = -6.18, p < .001^{***}$
9	4.55 (1.37)	3.51 (1.88)	$t(75) = 2.77, p < .004^{**}$	3.90 (1.80)	5.19 (1.47)	$t(68) = -3.24, p = .001^{***}$
10	4.13 (1.54)	2.56 (1.35)	$t(76) = 4.76, p < .001^{***}$	4.37 (1.75)	5.82 (1.50)	$t(75) = -3.92, p < .001^{***}$

\*  $\leq .05$ , \*\*  $\leq .01$ , \*\*\*  $\leq .001$

Table 6C2 Mean scores on manipulation checks for arguments from cause to effect with two causal connections (*SD* between parentheses; text numbers correspond to texts numbers in appendix 6B).

Text	Causal connection	Probability (1=very improbable; 7 = very probable)		Test result (all p-values one-tailed)
		Strong condition VI meets cause sufficiency criterion	Weak condition VII violates cause sufficiency criterion	
1	1	3.77 (1.38)	3.56 (1.59)	$t(68) = .58, p = .28$
	2	5.35 (1.17)	4.56 (1.48)	$t(68) = 2.43, p = .009^{**}$
2	1	4.59 (.91)	4.41 (1.45)	$t(64) = .66, p = .257$
	2	5.18 (1.23)	3.97 (1.46)	$t(76) = 3.94, p < .001^{***}$
3	1	4.77 (1.56)	3.45 (1.48)	$t(68) = 3.59, p < .001^{***}$
	2	5.79 (1.32)	4.48 (1.90)	$t(51.6) = 3.27, p = .001^{***}$
4	1	6.10 (1.17)	5.69 (1.24)	$t(76) = 1.51, p = .068$
	2	5.31 (1.00)	3.15 (1.53)	$t(65.6) = 7.35, p < .001^{***}$
5	1	5.44 (1.60)	5.28 (1.61)	$t(76) = .42, p = .337$
	2	5.21 (1.91)	3.49 (1.57)	$t(76) = 4.34, p < .001^{***}$
6	1	5.39 (1.36)	5.33 (1.03)	$t(68) = .19, p = .426$
	2	5.55 (1.36)	4.64 (1.31)	$t(68) = 2.83, p = .003^{**}$
7	1	3.85 (1.66)	4.97 (1.46)	$t(76) = -3.18, p = .001^{***}$ (contrary direction)
	2	4.44 (1.27)	3.56 (1.55)	$t(76) = 2.71, p = .004^{**}$
8	1	4.10 (1.35)	3.90 (1.81)	$t(68) = .53, p = .300$
	2	5.03 (1.53)	4.06 (1.61)	$t(68) = 2.55, p < .007^{**}$
9	1	5.72 (1.69)	6.49 (.94)	$t(59.7) = -2.49, p = .008^{**}$ (contrary direction)
	2	4.87 (1.61)	3.23 (1.95)	$t(76) = 4.05, p < .001^{***}$
10	1	3.59 (1.37)	3.38 (1.60)	$t(76) = .61, p = .273$
	2	5.08 (1.33)	3.51 (1.71)	$t(71.5) = 4.51, p < .001^{***}$

\*  $\leq .05$ , \*\*  $\leq .01$ , \*\*\*  $\leq .001$

Table 6C3 Mean scores on manipulation checks for arguments from cause to effect with a causal generalization (SD between parentheses; different letters refer to significant differences between groups appearing from post hoc tests (LSD),  $p < .05$ ; text numbers correspond to texts numbers in appendix 6B).

Text	Probability (1=very improbable; 7 = very probable)			Test result
	Condition (VIII) Strong: meets criteria (6)-(7) a	Condition (IX) Weak: violates relevant cause criterion (6) b	Condition (X) Weak: violates relevant effect criterion (7) a	
1	4.03 (1.20) a	4.79 (1.44) b	4.13 (1.49) a	$F(2, 114) = 3.56, p = .032^*$
2	3.82 (1.25) a	4.77 (1.23) b	3.00 (1.45) c	$F(2, 106) = 15.57, p = .000^{***}$
3	5.23 (.99)	5.62 (1.57)	4.90 (1.57)	$F(2, 114) = 2.57, p = .081$
4	4.67 (1.66) a	4.49 (1.43) ab	3.77 (1.56) b	$F(2, 106) = 3.09, p = .050^*$
5	4.90 (1.30)	4.92 (1.26)	5.00 (1.38)	$F(2, 106) = .06, p = .946$
6	4.38 (1.39)	4.28 (1.59)	3.67 (1.42)	$F(2, 114) = 2.73, p = .070$
7	3.38 (1.27) a	3.39 (1.36) a	4.08 (1.35) b	$F(2, 106) = 3.42, p = .036^*$
8	3.82 (1.62)	3.85 (1.55)	4.13 (1.40)	$F(2, 114) = .49, p = .615$
9	4.67 (1.28)	5.36 (1.35)	5.03 (1.22)	$F(2, 106) = 2.81, p = .065$
10	4.57 (1.30)	4.46 (1.50)	4.41 (1.82)	$F(2, 105) = .09, p = .919$

\*  $\leq .05$ , \*\*  $\leq .01$ , \*\*\*  $\leq .001$

Table 6C4 Mean scores on manipulation checks for arguments from example (*SD* between parentheses; a lower number indicates lower perceived typicality; different letters a-d refer to significant differences between groups,  $p < .05$ ).

Text	Condition (XI) typical example A1	Condition (XII) atypical example A1	Condition (XIII) typical example A2	Condition (XV) atypical example A2	Test result
1	5.54 (1.70) a	4.23 (1.94) b	6.03 (1.53) a	4.54 (2.06) b	$F(3,144) = 7.78, p = .000^{***}$
2	5.69 (1.54) a	4.05 (2.01) b	6.21 (1.26) a	2.71 (1.90) c	$F(3,144) = 30.83, p = .000^{***}$
3	5.82 (1.32) a	2.72 (1.82) b	5.81 (1.22) a	3.38 (1.91) b	$F(3,144) = 37.09, p = .000^{***}$
4	6.03 (1.05) a	3.05 (1.96) b	5.92 (1.42) a	4.69 (1.82) c	$F(3,144) = 27.06, p = .000^{***}$
5	5.69 (1.52) a	4.41 (2.09) b	4.46 (2.05) b	4.44 (2.00) b	$F(3,152) = 4.15, p = .007^{**}$
6	5.49 (1.85) a	3.00 (1.26) b	5.38 (1.50) a	4.10 (1.71) c	$F(3,144) = 18.41, p = .000^{***}$
7	6.23 (1.13) a	4.46 (1.94) b	5.85 (1.65) a	3.35 (1.85) c	$F(3,144) = 21.87, p = .000^{***}$
8	6.38 (.88) a	5.31 (1.61) b	5.77 (1.23) b	3.79 (1.91) c	$F(3,144) = 21.77, p = .000^{***}$
9	6.23 (.88) a	4.49 (1.83) b	5.90 (1.45) a	4.23 (2.03) b	$F(3,144) = 13.34, p = .000^{***}$
10	6.36 (1.22) a	4.18 (1.80) b	6.64 (.87) a	3.31 (1.82) c	$F(3,152) = 47.15, p = .000^{***}$

\*\*  $\leq .01$ , \*\*\*  $\leq .001$

## Summary in Dutch

**Titel:** The good, the bad and the persuasive

**Ondertitel:** Normatieve kwaliteit en daadwerkelijke overtuigingskracht van autoriteitsargumenten, argumenten van oorzaak naar gevolg en voorbeeldargumenten

### Inleiding

Veel onderzoekers hebben zich de vraag gesteld hoe mensen een standpunt evalueren dat hun in een persuasieve boodschap wordt voorgelegd. Zogenaamde *dual-process*-modellen, zoals Petty en Cacioppo's invloedrijke Elaboration Likelihood Model (ELM), veronderstellen dat mensen een standpunt op twee manieren kunnen evalueren: op basis van een zorgvuldige afweging van relevante argumenten ('centrale route') of op basis van niet-inhoudelijke kenmerken ('perifere route'), zoals de aantrekkelijkheid of status van de bron. Op welke manier de verwerking van een boodschap plaatsvindt, hangt af van een aantal factoren. Als mensen gemotiveerd en in staat zijn om de boodschap te verwerken, dan zijn ze eerder geneigd de argumenten aandachtig te beoordelen; anders richten ze zich eerder op perifere signalen.

Onder de voorwaarden dat men voldoende gemotiveerd is en in staat is om de boodschap te verwerken, kan argumentkwaliteit een belangrijke rol spelen in het overtuigingsproces. Sterke argumenten leiden tot acceptatie van het standpunt, terwijl zwakke argumenten niet tot overtuiging leiden. Gezien de vooraanstaande plaats die argumentkwaliteit inneemt in modellen van het overtuigingsproces, zou men verwachten dat onderzoekers een helder idee hebben van wat het begrip *argumentkwaliteit* inhoudt en dat ze voldoende inzicht hebben in cognitieve processen bij taalgebruikers die argumentkwaliteit beoordelen. Dit is echter niet het geval.

Voor het onderzoek naar persuasieve processen is het van belang om scherp te stellen wat een argument precies sterk of zwak maakt. Omdat argumentkwaliteit en de beoordeling ervan kan afhangen van het type argument dat is gebruikt, is het van belang bij onderzoek naar argumentkwaliteit rekening te houden met argumenttype. Een veel gebruikt argumenttype, zowel in het dagelijkse taalgebruik als in persuasieonderzoek, is het pragmatische argument. In pragmatische argumentatie wordt verdedigd dat een bepaalde handeling wenselijk of juist onwenselijk is op basis van de voordelen en/of nadelen van die handeling. Deze dissertatie richt zich

op drie verschillende argumenttypen, namelijk het autoriteitsargument, het argument van oorzaak naar gevolg en het voorbeeldargument. Deze argumenttypen kunnen worden gebruikt ter ondersteuning van één van de twee premissen van het pragmatische argument: de premisse die de waarschijnlijkheid uitdrukt dat er een bepaald effect zal optreden als gevolg van een handeling ('waarschijnlijkheidsclaim'). Daarom luidt het onderzoeksdoel:

Inzicht verschaffen in de specifieke kenmerken die de kwaliteit bepalen van het autoriteitsargument, het argument van oorzaak naar gevolg en het voorbeeldargument ter ondersteuning van de waarschijnlijkheidsclaim.

De normatieve argumentatietheorie biedt hier uitkomst. Ten eerste zijn daarin criteria geformuleerd, waaraan sterke argumenten zouden moeten voldoen. Er zijn bovendien criteria opgesteld die specifiek zijn voor bepaalde argumentatietypen. Ook aan het concretiseren van het centrale verwerkingsproces kan de normatieve argumentatietheorie een bijdrage leveren: op basis van de argumentatietypen en evaluatievragen die worden onderscheiden kunnen we voorspellen wat er in taalgebruikers om zou moeten gaan als ze argumenten aandachtig bestuderen en beoordelen. Deze hypothesen kunnen we vervolgens testen, om na te gaan of dat wat daadwerkelijk plaatsvindt in overeenstemming is met wat normatief vereist is in de argumentatietheorie. Dergelijk empirisch argumentatieonderzoek levert niet alleen een bijdrage aan het persuasieonderzoek, maar ook aan de argumentatietheorie, waarin men zich bezighoudt met de vraag of normen die op basis van theoretisch-analytische overwegingen zijn opgesteld ook deel uitmaken van de argumentatieve bagage van leken. De eerste onderzoeksvraag luidt daarom:

In welke mate corresponderen lekencriteria voor het autoriteitsargument, het argument van oorzaak naar gevolg en het voorbeeldargument met de criteria die zijn geformuleerd in de argumentatietheorie?

Daarbij zijn de volgende deelvragen geformuleerd:

- (a) Welke criteria zijn geformuleerd in de argumentatietheorie om de kwaliteit te beoordelen van het autoriteitsargument, het argument van oorzaak naar gevolg en het voorbeeldargument?
- (b) Welke criteria gebruiken leken om de kwaliteit te beoordelen van het autoriteitsargument, het argument van oorzaak naar gevolg en het voorbeeldargument?

Het ELM suggereert dat normatief sterke argumenten onder condities die centrale verwerking oproepen overtuigender zijn dan normatief zwakke varianten. Om die reden is de tweede onderzoeksvraag gesteld:

Zijn normatief sterke autoriteitsargumenten, argument van oorzaak naar gevolg en voorbeeldargumenten daadwerkelijk overtuigender dan normatief zwakke argumenten?

### **Deel I: Criteria voor argumentkwaliteit**

In het eerste deel van de dissertatie staat de vraag centraal in welke mate lekencriteria corresponderen met argumentatietheoretische criteria (onderzoeksvraag 1). In hoofdstuk 2 wordt deze vraag onderzocht voor het autoriteitsargument, in hoofdstuk 3 voor het argument van oorzaak naar gevolg en in hoofdstuk 4 voor het voorbeeldargument. Ieder onderzoek is min of meer op dezelfde wijze opgebouwd: eerst werd uitgezocht welke criteria er in de argumentatietheorie voor een bepaald argumenttype zijn geformuleerd (deelvraag 1a), vervolgens werd achterhaald welke criteria leken gebruiken om dat type te beoordelen (deelvraag 1b) en ten slotte werd nagegaan hoe de lekencriteria zich precies verhouden tot de criteria uit de argumentatietheorie en welke verschillen en overeenkomsten er op te merken zijn (onderzoeksvraag 1).

De argumentatietheoretische criteria zijn telkens onderzocht met literatuuronderzoek. Eerst werd een selectie van publicaties gemaakt in het domein van de argumentatieleer en vervolgens werden de evaluatiecriteria die in deze publicaties werden gevonden voor de onderzochte argumenttypen geïnventariseerd en gegroepeerd op basis van gemeenschappelijke kenmerken.

Om na te gaan wat de beste manier was om de lekencriteria te onderzoeken, werd besloten een methodologische voorstudie uit te voeren. De studie werd uitgevoerd binnen het onderzoek naar lekencriteria voor het autoriteitsargument (hoofdstuk 2). In deze studie werden respondenten door middel van casussen gestimuleerd na te denken over hun criteria voor argumentkwaliteit. In een gesloten casus kregen respondenten een standpunt voorgelegd en een lijst met zeven argumenten die elk niet aan verschillende, argumentatietheoretische criteria voldeden. De taak van de respondenten was een rangordening te maken van de argumenten, waarbij het sterkste argument op de eerste positie stond en het zwakste op de laatste positie. Nadat de respondenten een

rangordening hadden opgeschreven, werd hun om toelichting gevraagd.

Waar de gesloten casus bedoeld was om respondenten criteria te *ontlokken*, was de open casus bedoeld om na te gaan welke criteria leken *spontaan* genereren. In de open casus werd de respondenten een standpunt voorgelegd, maar bedachten ze zelf een sterk en een zwak argument ter ondersteuning van het standpunt. In een daaropvolgend gesprek werd de respondenten gevraagd om uit te leggen op basis waarvan het onderscheid tussen sterk en zwak gemaakt was. De casussen werden in twee verschillende omstandigheden voorgelegd: in een individueel interview of in een focusgroep waaraan zes respondenten deelnamen.

Op basis van deze methodologische voorstudie (N = 48) werd vastgesteld dat individuele interviews, waarin respondenten door middel van een open en een gesloten casus worden aangezet tot reflectie op argumentkwaliteit, de meeste lekencriteria boven water brengen. Om die reden werd in het onderzoek naar lekencriteria voor het argument van oorzaak naar gevolg (hoofdstuk 3) en het voorbeeldargument (hoofdstuk 4) verder gewerkt met deze methode.

Het antwoord op onderzoeksvraag 1 kan als volgt worden geformuleerd: zoals te verwachten valt op basis van de argumentietheorie, gebruiken leken schema-specifieke criteria om de kwaliteit te bepalen van het autoriteitsargument, het argument van oorzaak naar gevolg en het voorbeeldargument. Deze criteria corresponderen in belangrijke mate met de criteria die in de argumentatietheorie voor deze argumenttypen zijn opgesteld. Wel lijken sommige criteria breder gedragen te worden dan andere. Voor het autoriteitsargument geldt dat er relatief hoge overeenstemming bestaat onder respondenten over het relevante- expertise-criterium ('the relevant expertise criterion'), het expertcriterium ('the expert criterion'), het betrouwbaarheids criterium ('the trustworthiness criterion'), het criterium dat verwijst naar het vermogen om bewijs te verschaffen ('the ability to provide evidence criterion'), het recentheids criterium ('the recency criterion') en het slag-om-de-arm-criterium ('the hedged statement criterion'). De eerste vijf criteria zijn lekencriteria die met de argumentatietheorie corresponderen; het laatstgenoemde criterium stemt niet overeen met de argumentatietheorie. Voor het argument van oorzaak naar gevolg geldt dat er relatief hoge overeenstemming bestaat over het criterium dat verwijst naar de oorzaak als voldoende voorwaarde ('the cause sufficiency criterion'). Dit leken criterium correspondeert met de argumentatietheorie. Voor het voorbeeldargument gaat op dat er



flinke overeenstemming bestaat over het relevantie criterium ('the relevance criterion'), het aantal-voorbeelden-criterium ('the number of examples criterion'), het correctheid-van-het-voorbeeld-criterium ('the example accuracy criterion') en het causaliteitscriterium ('the causality criterion'). De eerste drie criteria corresponderen met de argumentatietheorie; het laatste criterium niet. Met betrekking tot de lekencriteria waarover aanzienlijke overeenstemming bestaat, kan voorzichtig geconcludeerd worden dat deze doorgaans deel uitmaken van de argumentatieve vaardigheden waarover kritische taalgebruikers kunnen beschikken.

Sommige criteria die in de argumentatietheorie zijn geformuleerd worden niet door leken gebruikt, maar in de meeste gevallen kan dat worden verklaard door het materiaal, dat respondenten geen enkele reden gaf om deze criteria te gebruiken. Werden irrelevante criteria gebruikt door leken, dat wil zeggen criteria die eigenlijk verbonden zijn met een ander argumentschema, dan was het gebruik van deze irrelevante criteria begrijpelijk op basis van het feit dat het in het materiaal ging om voorgenomen acties en de effecten ervan.

## **Deel II: De relatie tussen argumentkwaliteit en daadwerkelijke overtuigingskracht**

In het tweede deel van de dissertatie staat de vraag centraal of normatief sterke autoriteitsargumenten, argumenten van oorzaak naar gevolg en voorbeeldargumenten daadwerkelijk overtuigender zijn dan normatief zwakke (onderzoeksvraag 2). Deze vraag is onderzocht met behulp van experimenteel onderzoek. Eerst werd een verkennend experimenteel onderzoek uitgevoerd (hoofdstuk 5), om uit te zoeken in hoeverre er een relatie is tussen de daadwerkelijke overtuigingskracht en de waargenomen redelijkheid van argumenten. In het experiment kregen respondenten een vragenlijst voorgelegd met 16 korte tekstjes. Ieder tekstje bestond uit een argument en een conclusie die door dat argument ondersteund werd. Het argument was een autoriteitsargument of een argument van oorzaak naar gevolg en was normatief sterk of normatief zwak. De daadwerkelijke overtuigingskracht werd gemeten door respondenten (N = 50) te laten aangeven op een zevenpunts-Likert-schaal in hoeverre ze het eens waren met de laatste zin van het tekstje (de conclusie). Waargenomen redelijkheid werd gemeten door een andere groep respondenten (N = 50) op een zevenpunts-Likert-schaal te laten

aangeven in hoeverre ze het argument een goede ondersteuning vonden voor de conclusie in de laatste zin.

De voorstudie toonde aan dat de daadwerkelijke overtuigingskracht en waargenomen redelijkheid met elkaar verbonden zijn: argumentatie die redelijker wordt gevonden is over het algemeen ook overtuigender in de situatie waarin de argumentatie zorgvuldig wordt bestudeerd. Verder bleek uit de resultaten dat normatief sterke argumenten van oorzaak naar gevolg redelijker gevonden worden en ook overtuigender zijn dan normatief zwakke argumenten van oorzaak naar gevolg. Voor autoriteitsargumenten hangt het effect van argumentkwaliteit af van het argumentatieschema dat is gebruikt en ook van het evaluatiecriterium dat is gebruikt om het argument normatief sterk of zwak te maken. In specifieke termen: wanneer een descriptieve claim wordt verdedigd, krijgt een ondersteunend argument met een onbetrouwbare bron minder positieve oordelen dan een ondersteunend argument met een betrouwbare bron, maar de relevantie van de expertise maakt geen verschil. Wanneer een normatieve claim wordt verdedigd, kunnen sterke autoriteitsargumenten op hogere scores rekenen dan zwakke autoriteitsargumenten en is er geen interactie-effect tussen criterium en argumentkwaliteit, zoals voor het schema met een descriptieve claim wel gevonden is.

Al met al werd geconcludeerd uit het verkennende experimentele onderzoek dat de resultaten voor de daadwerkelijke overtuigingskracht en die voor de waargenomen redelijkheid een vergelijkbaar patroon laten zien. Er werd dan ook besloten in het daarop volgende hoofdexperiment alleen de daadwerkelijke overtuigingskracht als afhankelijke variabele te meten.

In het hoofdexperiment (hoofdstuk 6) kregen respondenten een vragenlijst voorgelegd. In de vragenlijst stonden 34 tekstjes. Ieder tekstje bevatte een argument en een conclusie die door het argument werd ondersteund. De gebruikte argumentatie was autoriteitsargumentatie, argumentatie van oorzaak naar gevolg, voorbeeldargumentatie of cirkelargumentatie. Cirkelargumentatie werd in het materiaal aangeboden om achteraf te kunnen vaststellen of respondenten de argumentatie met voldoende aandacht hadden beoordeeld. In het geval van de drie andere typen argumentatie was de argumentatie normatief sterk of normatief zwak. De daadwerkelijke overtuigingskracht werd gemeten door respondenten (N = 200) te laten aangeven op een zevenpunts-Likert-schaal in hoeverre ze het eens waren met de conclusie in de laatste zin van het tekstje.

De resultaten lieten zien dat voor het argument van oorzaak naar gevolg en het voorbeeldargument in het algemeen geldt dat normatief sterke argumenten daadwerkelijk overtuigender zijn dan normatief zwakke argumenten. Voor het autoriteitsargument geldt dat het normatief sterke autoriteitsargument wel overtuigender is dan een argument op basis van een onbetrouwbare bron, maar even sterk is als een argument op basis van een bron die (a) minder waarschijnlijk is als expert, of (b) een minder relevant expertisegebied heeft. Het verschil tussen de sterke variant en de zwakke variant waarin de bron een mening heeft die minder aansluit bij de conclusie, bereikte net niet het conventionele significantieniveau.

Uit het onderzoek in deel II werd al met al geconcludeerd dat normatief sterke argumenten niet per se daadwerkelijk overtuigender zijn dan normatief zwakke argumenten; dat hangt af van het betreffende argumenttype en het specifieke evaluatiecriterium dat wordt geschonden in het normatief zwakke argument. De voorspelling die voortvloeit uit het ELM, namelijk dat normatief sterke argumenten eerder tot overtuiging leiden dan normatief zwakke argumenten, wordt dan ook niet volledig ondersteund door de resultaten van dit onderzoek.

### **Algemene conclusie en discussie**

Hoofdstuk 7 blikt terug op de resultaten in deel I en II van de dissertatie. Als die resultaten aan elkaar gerelateerd worden, dan wordt het volgende geconcludeerd: ook al gebruiken leken bepaalde criteria om de kwaliteit van autoriteitsargumenten vast te stellen, de overtuigingskracht van autoriteitsargumenten die deze evaluatiecriteria schenden is er niet altijd minder om. Deze bevinding strookt niet met de hypothese dat normatief sterke argumenten bij centrale verwerking overtuigender zijn dan normatief zwakke argumenten, iets dat men op basis van het ELM wel zou verwachten.

Verklaringen kunnen ten eerste worden gezocht in het verschil tussen de taak die respondenten in het kwalitatieve onderzoek uitvoerden en de taak die respondenten in het kwantitatieve, experimentele onderzoek uitvoerden. De taak in het kwalitatieve onderzoek, waarbij argumenten vergeleken en gerangordend werden, stimuleerde respondenten waarschijnlijk veel aandacht te besteden aan de kwaliteit van het argument. In het experiment werd respondenten gevraagd in hoeverre ze het eens of oneens waren met de conclusie, dus is het maar de vraag of ze geneigd waren veel aandacht aan de argumentkwaliteit te schenken. In die situatie waren

sommige experimentele manipulaties wellicht te subtiel om een effect te vinden op de daadwerkelijke overtuigingskracht.

Als alternatieve verklaringen worden genoemd: leken vinden sommige criteria wellicht belangrijker dan andere criteria, ook al zijn ze van alle criteria op de hoogte, of ze zijn niet in staat alle criteria die in het geheugen opgeslagen zijn daadwerkelijk toe te passen tijdens de taak.

In hoofdstuk 7 komen ook beperkingen van de uitgevoerde studies aan de orde, die betrekking hebben op kwesties van validiteit en van betrouwbaarheid. Zo zijn er vraagtekens bij het ontwerp van de open casussen en het concrete niveau waarop leken hun evaluatiecriteria verwoordden (validiteit) en bij het gegeven dat slechts een stap in de kwalitatieve analyse is getest op overeenstemming met een tweede beoordelaar (betrouwbaarheid). Tegen de experimentele studies kan worden ingebracht dat er geen controleconditie is gebruikt en dat de claims te neutraal waren om voldoende betrokkenheid van respondenten op te roepen. Een andere kanttekening bij de experimenten is dat korte tekstjes zijn gebruikt, waardoor de manipulaties mogelijk opvallender waren dan wanneer langere, meer alledaagse betogen zijn gebruikt.

Vanwege deze beperkingen is vervolgonderzoek aan te bevelen, waarbij vooral aandacht komt voor het ontwerp van materiaal dat bij respondenten voldoende betrokkenheid teweegbrengt en tegelijkertijd claims bevat die bij respondenten nog ter discussie staan. Daarnaast is het de moeite waard te onderzoeken wat er nog van de gevonden effecten overblijft, wanneer manipulaties van argumentkwaliteit in langere teksten worden verwerkt.

## Curriculum vitae

Ester Šorm werd op 3 februari 1981 in Hilversum (Nederland) geboren. In 1998 haalde ze aan het Gemeentelijk Gymnasium in Hilversum haar vwo-diploma. Vervolgens studeerde ze Psychologie (1998-2002) en Taal- en Cultuurstudies (1998-2004, cum laude) aan de Universiteit Utrecht. In de jaren 2005-2009 werkte ze als promovendus bij de afdeling Bedrijfscommunicatie van de Radboud Universiteit Nijmegen. Ze participeerde in een door de Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) gesubsidieerd project, getiteld 'The quality of pragmatic arguments'. Momenteel is ze postdoctoraal onderzoeker bij de afdeling Taal en Communicatie van de Vrije Universiteit Amsterdam. Daar werkt ze aan het project 'Identifying and describing visual metaphor', dat gefinancierd wordt door NWO.

---

Ester Šorm was born in Hilversum (the Netherlands) on 3 February 1981. After graduating from the Gemeentelijk Gymnasium in Hilversum in 1998, she studied Psychology (1998-2002) and Language and Culture Studies (1998-2004, cum laude) at Utrecht University. From 2005 to 2009, she worked as a PhD-student at the Department of Business Communication Studies of the Radboud University Nijmegen. She participated in a project titled 'The quality of pragmatic arguments', which was funded by the Netherlands Organisation for Scientific Research (NWO). She currently is a postdoctoral researcher in the Department of Language and Communication at VU University Amsterdam, where she works on the project 'Identifying and describing visual metaphor', financed by NWO.